JOURNAL OF Experimental Child Psychology

EDITOR: Sidney W. Bijou

ASSOCIATE EDITORS: Richard S. Bogartz
David Zeaman

EDITORIAL BOARD:

Donald M. Baer
Albert Bandura
Wesley C. Becker
Harry Beilin
Gordon N. Cantor
Peter Eimas
Robert L. Fantz
Eleanor J. Gibson
Eugene S. Gollin
Willard W. Hartup
Raymond H. Hohle
Herbert Kaye
Tracy S. Kendler
Helen L. Koch
Eugene R. Long
David S. Palermo
Herbert L. Pick
Hayne W. Reese
Bryan E. Shepp
Murray Sidman
Einar Siqueland
Patrick Suppes
Joachim Wohlwill

Volume 7, Number 1, February 1969

ACADEMIC PRESS
New York and London

CONTENTS

CATHERINE E. SNOW AND M. SAM RABINOVITCH. Conjunctive and Disjunctive Thinking in Children 1

BARBARA MORGAN WILCOX. Visual Preferences of Human Infants for Representations of the Human Face 10

ANNE SARAVO AND MAY KOLODNY. Learning Set and Shift Behavior in Children 21

ROY GLADSTONE. Age, Cognitive Control and Extinction 31

W. C. SHEPPARD. Operant Control of Infant Vocal and Motor Behavior . . 36

MARSHALL M. HAITH, WILLIAM KESSEN, AND DORIS COLLINS. Response of the Human Infant to Level of Complexity of Intermittent Visual Movement 52

MICHAEL LEWIS AND SUSAN GOLDBERG. The Acquisition and Violation of Expectancy: An Experimental Paradigm 70

PAUL WEISBERG. Operant Procedures for the Establishment of Stimulus Control in Two-Year-Old Infants 81

JOSEPH E. SPRADLIN, DEAN L. FIXSEN, AND FREDERIC L. GIRARBEAU. Reinstatement of an Operant Response by the Delivery of Reinforcement during Extinction 96

ROBERT G. WAHLER. Infant Social Development: Some Experimental Analyses of an Infant-Mother Interaction during the First Year of Life . . . 101

JOHN HAMILTON AND JON STANDAHL. Suppression of Stereotyped Screaming Behavior in a Profoundly Retarded Institutionalized Female 114

RACHEL KEEN CLIFTON AND WILLIAM J. MEYERS. The Heart-Rate Response of Four-Month-Old Infants to Auditory Stimuli 122

N. H. PRONKO. Biotelemetry in Child Study 136

KEITH G. SCOTT. The Design of Mobile Laboratories for Behavioral Research with Children 143

KENNEDY T. HILL AND BARBARA E. MOELY. Social Reinforcement as a Function of Task Instructions, Sex of *S*, Age of *S*, and Baseline Performance . . 153

Published bimonthly at Exchange Place, Hanover, Pa. 17731 by Academic Press, Inc., 111 Fifth Avenue, New York, N. Y., 10003.

In 1969, Volumes 7 and 8 (consisting of 3 issues each) will be published.

Price: $22.00

Private subscriptions (for subscriber's personal use): $10.00

All correspondence and subscription orders should be sent to the office of the Publishers at 111 Fifth Ave., New York, N.Y. 10003

Second-class postage paid at Hanover, Pennsylvania 17331

© 1969 by Academic Press, Inc.

32

Additional Volume of *Journal of Experimental Child Psychology* to be published in 1969

In order to permit more rapid publication of a number of important manuscripts now on hand, an additional volume will be released during 1969.

The revised publication schedule is as follows:

Volume 7—February, April, June
Volume 8—August, October, December

The price of each 1969 volume is $22.00 for institutional subscribers; and $10.00 for private subscribers certifying that the journal is for personal use.

Academic Press, Inc.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 1–9 (1969)

LIBRARY

Conjunctive and Disjunctive Thinking in Children[1]

CATHERINE E. SNOW AND M. SAM RABINOVITCH
McGill University

In light of the finding (Bruner *et al.*, 1956) that disjunctive concepts are more difficult for adults than conjunctive concepts, a developmental study of the relative difficulty of conjunctive and disjunctive concepts was carried out. Five groups of children, ranging in age from 5 years to 13 years, were tested on logically equivalent conjunctive and disjunctive concepts. It was found that children of all ages make more errors on disjunctive than conjunctive tasks. The strategies used for conjunctive tasks were more appropriate at all age levels. Because of the fact that the difference in difficulty is quite constant over the age groups, it was concluded that some aspect of conjunctive groupings is more natural. As a result, they occur more frequently and there is greater opportunity to learn to reason correctly about them.

It has been reported that normal adult *Ss* have much more difficulty learning disjunctive concepts than conjunctive concepts of equivalent logical difficulty (Bruner, Goodnow, and Austin, 1956). At the same time it was observed that the strategies used in disjunctive tasks were for the most part completely inappropriate, and were in fact the very same strategies that are correctly used with conjunctive tasks. Adults have a bias toward thinking conjunctively, even when they know they are not dealing with conjunctive problems. Conjunctive concepts are defined as those where positive instances display all the values named by the concept. Positive instances of disjunctive concepts, on the other hand, need display only one of the values named by the concept. Two different hypotheses have been suggested to account for the greater difficulty of disjunctive concepts. The first is that conjunctive concepts are more common in everyone's thinking and are easier to talk about, and so strategies for thinking about them are more practiced and become better learned. The second hypothesis is that disjunctive concepts are inherently more difficult, or psychologically complex, because the human brain is not built to think disjunctively.

[1] This report is based on an M.A. thesis submitted to the Department of Psychology, McGill University, by the senior author, a National Science Foundation graduate fellow.

We express our gratitude to St. George's School of Montreal and the Westmount Y.M.C.A. for their help in making subjects available.

© 1969 by Academic Press, Inc.

Inst. of Ment. & Psyl. Research
(I.R.T.) 3879

TABLE 1
TESTING MATERIALS

Attributes	Values	Concept	Relevant attributes of positive instances	Relevant attributes of negative instances	
Card color Figure color Figure shape Figure size Number of figures	Yellow or white Red or blue Square or circle Large or small One or two	Red and square (conjunctive)	Red and square (eight cards)	Blue and square red and circle (four cards each)	Deck 1-a (16 cards)
		Red or square (disjunctive)	Blue and square red and circle (four cards each)	Blue and circle (eight cards)	Deck 1-b (16 cards)
Card color Figure color Figure size Figure shape	Yellow or white Red or blue Large or small Square, circle or triangle	Large and white (conjunctive)	Large and white (eight cards)	Small and white large and yellow (four cards each)	Deck 2-a (12 cards)
		Large or white (disjunctive)	Small and white large and yellow (four cards each)	Small and yellow (eight cards)	Deck 2-b (12 cards)

This was done only to direct the child's attention to the task of finding a rule. If the choice was incorrect, the *E* told the *S* he was wrong, and asked if he knew why. Then the card was replaced in its correct pile before the *S* went on. The *S*s were allowed to look through previous choices whenever they wished. After the *S* was finished with each deck of cards, he was asked to state the correct rule, if he could.

The *E* recorded the order in which the cards were chosen, the errors, and the reason given. An attempt to measure how competently the *S* was approaching the task was made by specially noting the first card chosen. The ideal choice, in view of the *S*'s two aims of avoiding mistakes and discovering the rule, is to choose a card which differs from the illustrative card (chosen by the *E*) in only one attribute, and to place it in the same group as the illustrative card. A card different from the illustrative card in only one attribute out of a possible four or five is most likely to belong in the same group as the illustrative card. This choice, then, is least likely to be a mistake. If it is a mistake, however, it is still valuable because it then provides the information that the changed attribute is relevant, and the *S*, if he uses this information, knows half the rule.

RESULTS

The principle measure taken was the number of errors committed during the task, that is, the number of cards originally placed in the incorrect group. The errors for the two disjunctive tasks were summed, as were the errors for the two conjunctive tasks, to obtain one score for each *S* on each task. An analysis of variance was performed on the scores. The results of this test are given in Table 2.

It can be seen from Table 2 that the difference between the conjunctive and disjunctive tasks is significant, and that the age difference is significant. The lack of interaction between age and task indicates that the disjunctive task is harder at all ages, and thus fails to support the hypothesis that the disjunctive task becomes increasingly harder with age.

The magnitude of the differences can be seen in Fig. 1. It is quite striking that the differences in difficulty between conjunctive and disjunctive tasks are almost constant for all ages, except for the convergence at the youngest age group. Since there is no significant age by task interaction, it would be risky at this stage to think of this convergence as anything other than random variation.

A second measure, a reflection of the logic used in approaching the two tasks, is the appropriateness or inappropriateness of the initial choice of cards in each task by each *S*. If the ideal initial choice is made

TABLE 2
ANALYSIS OF VARIANCE OF ERROR SCORES

Source of variation	SS	*df*	MS	F	
Between *Ss*					
Age	332.98	4	83.25	10.30	$p < .01$
Sex	.37	1	.37	—	
Age × sex	36.65	4	9.16	—	
Ss within groups	703.01	87	8.08		
Within *Ss*					
Task	79.26	1	79.26	15.33	$p < .01$
Age × task	16.01	4	4.00	—	
Sex × task	.53	1	.53	—	
Age × sex × task	4.42	4	1.10	—	
Task × *Ss* within groups	449.78	87	5.17		

in both conjunctive tasks and in neither disjunctive task, the *S* is scored in the C ≫ D category. If the *S* chooses the ideal initial card in the conjunctive task only once, and in the disjunctive task never, or if he chooses the ideal initial card in both conjunctive tasks and in only one disjunctive task, then he is placed in the C > D category. If he performs equally well in both tasks, or if he chooses the ideal card more often in the disjunctive tasks, then he is placed in the C ⩽ D category. The resulting distribution can be seen in Table 3. A χ^2 performed on

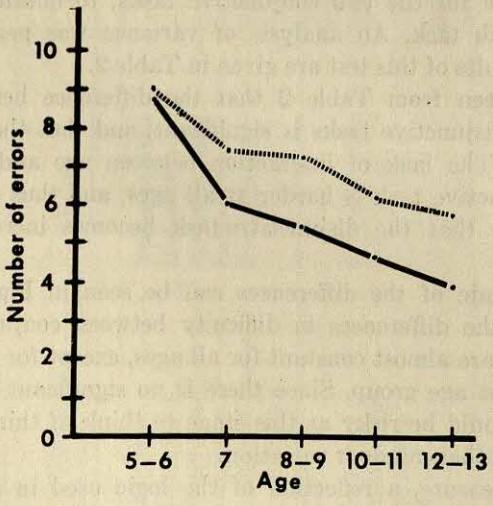


FIG. 1. Number of errors as a function of age. ——— conjunctive task; - - - - - - - - - disjunctive task.

these data is significant at the .05 level. This indicates that the different age groups are differentially able to use the ideal approach to solving conjunctive and disjunctive problems. Specifically, the eight- and eleven-

TABLE 3
FREQUENCY OF IDEAL RESPONSE IN CONJUNCTIVE AND DISJUNCTIVE TASKS

Age	5	7	8	11	13
$C \gg D$	5	5	11	11	6
$C > D$	7	6	7	7	3
$C \leq D$	9	9	2	2	7

year-olds perform much better on conjunctive tasks, while the five-, seven- and thirteen-year-olds perform only slightly better on conjunctive tasks.

Another reflection of the logic used in approaching the tasks is whether the initial choice is a positive instance or a negative instance. The ideal choice for conjunctive tasks is positive, for disjunctive tasks, negative. The figures for both tasks are given in Table 4. Two inde-

TABLE 4
FREQUENCY OF POSITIVE AND NEGATIVE INSTANCES IN CONJUNCTIVE AND DISJUNCTIVE TASKS

Age	5	7	8	11	13
Conjunctive task					
pos (ideal)	37	37	38	39	29
neg	5	3	2	1	3
Disjunctive task					
pos	17	20	29	29	16
neg (ideal)	25	20	11	11	16

pendent χ^2 tests were performed on these figures. The first, on the frequencies of positive and negative choices in conjunctive tasks, is non-significant. The second, for disjunctive tasks, is significant at the .01 level. Thus there are age differences for disjunctive but not for conjunctive tasks. All age groups choose primarily positive instances when solving conjunctive tasks, but only eight- and eleven-year-olds choose primarily positive instances when solving disjunctive tasks, the other groups choosing about half negative instances.

DISCUSSION

There is no support for the hypothesis that disjunctive tasks are harder for adults simply because there is more opportunity to learn and use con-

junctive reasoning. Either disjunctive reasoning really is harder for children of all ages, or else the youngest children tested in this experiment had already been exposed to sufficient learning experiences which favored conjunctive thought. Assuming the learning hypothesis, however, one would still predict that some learning would go on after 5 years of age, and that the curves would continue to diverge. This does not happen.

Analysis of the strategies used indicates that children of different ages do not perform equally well in choosing strategies for disjunctive tasks. Eight- and eleven-year-olds perform relatively worse than either older or younger children on both measures indicating strategy. It is possible that the younger children are merely performing randomly, without applying any consistent strategy, and that as soon as they become mature enough to use strategies, they use conjunctive ones with all the tasks given to them. Or it is possible that younger children are approaching the disjunctive tasks without the disadvantage of having learned bad strategies, but that they cannot yet use the information they acquire, so the disjunctive task remains very hard for them.

Considering that number of errors is the more reliable measure of ability, the overall conclusion must be that disjunctive tasks are more difficult for children just as they are for adults. This conclusion is supported by the finding (King, 1966) that disjunctive rules are harder to learn than conjunctive rules for children as well as adults, in a situation where the *Ss* have been told the relevant attributes. However, the final assurance that differential learning is not the main factor in the greater difficulty of disjunctive tasks will come from a study of children younger than those used in the present experiment or the one reported by King (1966).

The question of why disjunctive tasks are consistently more difficult cannot be answered, but deserves careful thought. Perhaps disjunctive concepts are less useful because they are associated with a lower level of certainty than conjunctive concepts. A disjunctive grouping does not give one any absolute knowledge about the members of the group, so there is, in a sense, no point in thinking that way. Or perhaps disjunctive ideas are more complex in terms of brain processes, so the brain automatically works in the simpler conjunctive way. This would imply, for instance, that "red *and* square" requires only one step in the thought process while "red *or* square" requires two.

The results of this experiment lead to the conclusion that there is some aspect of disjunctive groupings that make them inherently more complex than conjunctive groupings. As a result, conjunctive groupings are preferred and occur more frequently, providing greater opportunity to learn to reason about them.

REFERENCES

BRUNER, J. S., GOODNOW, J. J., AND AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

BRUNER, J. S., OLVER, R. R., GREENFIELD, P. R., et al. *Studies in cognitive growth.* New York: Wiley, 1966.

HOVLAND, C. L., AND WEISS, W. Transmission of information concerning concepts through positive and negative instances. *Journal of Experimental Psychology,* 1953, **45**, 175–182.

HUNT, E. B., AND HOVLAND, C. L. Order of consideration of different types of concepts. *Journal of Experimental Psychology,* 1960, **59**, 220–225.

KING, W. L. Learning and utilization of conjunctive and disjunctive classification rules: A developmental study. *Journal of Experimental Child Psychology,* 1966, **4**, 217–231.

PIAGET, J. *The origins of intelligence in children.* New York: Universities Press, Inc., 1952.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 10–20 (1969)

Visual Preferences of Human Infants for Representations of the Human Face[1]

BARBARA MORGAN WILCOX[2]
Brown University

Thirty infants, 4, 10, and 16 weeks old viewed 7 pictures of faces: a photograph; a realistic drawing; and schematic drawings of a complete, 3 incomplete, and a scrambled face. Analyses of total fixation time and of duration of first fixation gave somewhat discrepant evidence about preferences among the stimuli. The data were interpreted as showing: no preference for the schematic over the symmetrically scrambled face, no difference between a face with no eyes and one with no nose and mouth, preference for more complex over less complex stimuli with complexity defined as the number of elements, and developmental changes in preferences among complete, unaltered faces. Generally, complexity outweighed "faceness" as a determiner of preferences.

Ever since Fantz (1961) reported that infants from birth to 6 months of age fixate a schematic drawing of a face longer than a scrambled version of the same drawing, it has been assumed that any stimulus containing the configuration of the facial features will be preferred by infants to any other stimulus equal in complexity. However, the differences in fixation time which Fantz found between these two stimuli were very small, and he performed no statistical tests. Also, no subsequent study in which a representation of a face has been compared with a scrambled version of the same stimulus has found a significant difference between the two stimuli across the entire age range tested (Fantz, 1966; Kagan, Henker, Hen-Tov, Levine, and Lewis, 1966; Lewis, Campbell, Bartels, and Fadel, 1966). However, an occasional instance of a statistically significant preference for the regular face over a scrambled or up-side down control stimulus has been found at isolated

[1] This paper is based on a thesis submitted in partial fulfillment of the requirements for the Ph.D. degree to Brown University, June, 1967. The research was conducted while the author was an NSF Graduate Fellow. The author gratefully acknowledges the assistance of Prof. Frances L. Clayton, who served as thesis advisor. Thanks are also due to Karen Lambe and Kathleen Decker for their help with preparation of the stimuli.

[2] Now at the Department of Psychology, Bishop College, 3837 Simpson-Stuart Rd., Dallas, Texas 75241.

but inconsistent age levels (Fantz, 1966, 2 to 3 months; Lewis *et al.*, 24-week-old girls; Watson, 1966, 14 weeks), and the supposition that representations of faces are generally preferred to other kinds of patterns has persisted.

There has been little consistency among the above studies on the particular stimuli used, and the ambiguous results obtained suggest that factors other than faceness, or degree of resemblance to a real face, may be affecting the infants' visual fixations. Thomas (1965) points out that the features in the scrambled-face control stimuli have not been symmetrically arranged, and that these control stimuli might therefore be considered to be more complex than the regular face, which is symmetrical about the vertical axis. Brennan, Ames, and Moore (1966) have demonstrated that complexity, interacting with age, can affect infants' visual preferences. This fact indicates a need for a direct investigation of the influence of symmetry and other aspects of complexity on infants' visual preferences among representations of the face.

A second factor which has varied considerably among studies of visual preferences and which has been shown to be of influence is the measure of fixation time which is used in the analysis of the data. Lewis, Kagan, and Kalafat (1966) report that the duration of the first fixation in each stimulus presentation is more sensitive to stimulus differences than is total fixation time cumulated across the entire stimulus presentation. Kagan and Lewis (1965) report that increasing the number of trials included in the data analysis can decrease the sensitivity to stimulus differences. These aspects of the response measure have not been consistent across the various studies of preferences for faces, and discrepant results might be partially attributed to these differences.

The experiment described in this paper was an investigation of factors which control infants' visual preferences among representations of the face. A number of different criteria have been used to define complexity in studies of infant visual preferences (Berlyne, 1958; Brennan, Ames, and Moore, 1966; Hershenson, Munsinger, and Kessen, 1965; and Thomas, 1965). Two of these definitions, symmetry and number of elements, were used in the present study. Four different response measures were analyzed in order to observe the effect that choice of response measure might have on the results.

METHOD

Stimuli. Seven black and white two-dimensional stimuli were constructed (Fig. 1). Five of these were based on the schematic face which Fantz (1961, 1966) used. A symmetrically scrambled control stimulus (Srambled) provided a control for complexity. This stimulus was formed

from the schematic face (Schematic) by rearranging the facial features into a non-facelike configuration symmetrical about the vertical axis. A comparsion of fixation times of this symmetrically scrambled face with the schematic face should provide a direct test of the importance of a facelike configuration as a determiner of visual preferences. Deletions of various features from the schematic face provided a set of stimuli which varied in number of elements. The schematic face and the scrambled face were considered to be more complex than the two in-

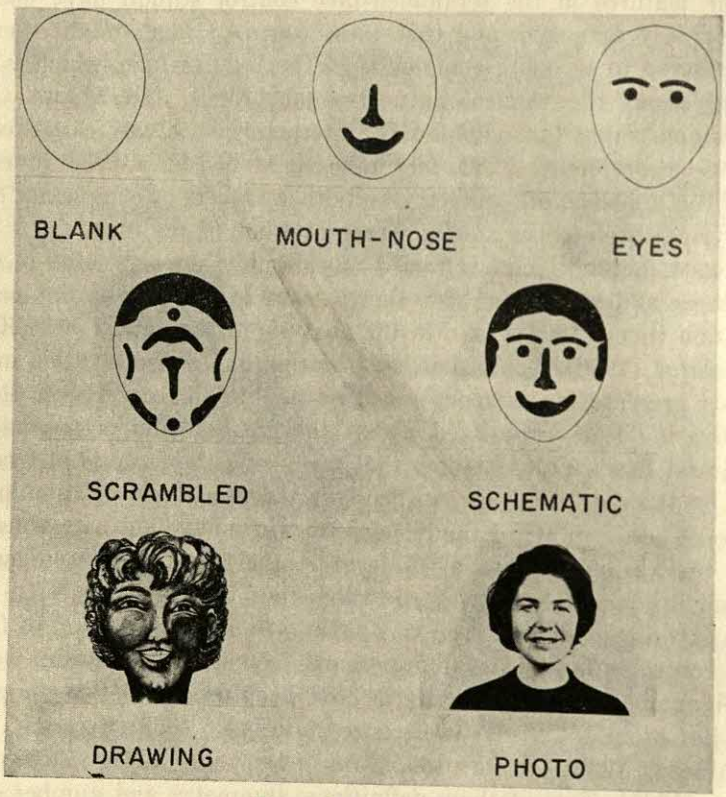


FIG. 1. The stimuli.

complete faces, which contained only the eyes (Eyes) and only the mouth and nose (Mouth-Nose); the two incomplete faces were considered to be more complex than the blank oval outline (Blank). A comparison between the two incomplete faces (Eyes and Mouth-Nose) was included to determine if some features of the face might be preferred to others. Three complete, unaltered faces were designed to vary in the degree of realism with which they represented the face. These three stimuli were the schematic face, a more realistic drawing of a woman's

face which contained continuous tones of gray and shaded areas (Drawing), and a photograph of a women's face (Photo). The drawing was a modification of the drawing of a boy's face which Ahrens (1954) used in his study of infant smiling.

The questions to be asked of the data required that comparisons be made among the fixation times for the seven stimuli. Because non-independent multiple *t* tests result in unknown reductions in significance level, a system of planned comparisons (Hays, 1963) was used whereby only selective testing was conducted. The statistical comparisons performed on these data are presented in Table 1. The use of planned comparisons provides for mutual independence among tests by establishing a sequence of comparisons wherein any set composed of all of the stimuli on both sides of one comparison must appear in its entirety on only one side of any subsequent comparison which is to include any member of the set. Thus, groupings of the stimuli for the statistical tests were pre-established according to the above rule for planned comparisons and based upon the primary questions which the study sought to answer.

TABLE 1
PLANNED COMPARISONS

1. Is there any difference between Eyes and Mouth-Nose?
   Comparison I: Eyes versus Mouth-Nose
2. Is Schematic preferred to Scrambled?
   Comparison II: Schematic versus Scrambled
3. Are there visual preferences among the three complete, unaltered faces?
   Comparison III: Photo versus Scrambled and Schematic
   Comparison IV: Drawing versus Scrambled, Schematic, and Photo
4. Are there visual preferences along the complexity continuum?
   Comparison V: Eyes and Mouth-Nose versus Scrambled, Schematic, Drawing, and Photo
   Comparison VI: Blank versus Eyes, Mouth-Nose, Scrambled, Schematic, Drawing, and Photo

Subjects. Three age groups of infants were used, with 10 *Ss* in each group. The ages sampled were 4 weeks ± 1 week (6 boys and 4 girls), 10 weeks ± 1 week (5 boys and 5 girls), and 16 weeks ± 1 week (6 boys and 4 girls). The *Ss* were patients of three pediatricians,[3] and the only criteria for selection were the infant's age and the mother's willingness to cooperate. For *Ss* who cried or fell asleep before a minimum number of stimuli had been presented the data were rejected. There were 17 such

[3] Jay M. Orson, M.D., Alfred Toselli, M.D., and Wilson F. Utter, M.D. generously supplied names of potential subjects from their patient files.

rejections in the 4-week-old group, four in the 10-week-old group, and six in the 16-week-old group.

Apparatus. The infant lay in a supine position in a cradle. Slides of the stimuli were projected onto a rear-projection screen located above and behind the infant's head. This screen was visible to the infant in a mirror placed directly over the infant's head and situated at an angle of 45° with the horizontal. The screen was at an optical distance of 26½ inches from the infant's eyes and the projected faces were 6½ inches high (14° of visual angle). According to the data on infants' ability to accommodate (Haynes, White, and Held, 1965), a viewing distance of 26½ inches would not be optimal for infants as young as four weeks. However, the fact that visual preferences were obtained in this age group indicates that the infants were able to discriminate among these stimuli at this distance. The cradle, mirror, and screen were all contained within an enclosure which had walls made of fine, white cheesecloth. A peephole centered between the mirror and the projection screen permitted observation of the infant's eyes.

Recording of responses. An observer watched the infant's eyes through the peephole and judged fixation of the stimulus when the corneal reflection of the patch of light from the stimulus was centered over the pupil of the infant's eye. An interobserver reliability of .91 (Pearson product moment correlation coefficient, $p < .01$) was obtained with two observers recording simultaneously. If the two eyes were not converging on the stimulus, as occurred frequently with four-week-old *Ss*, then fixations were judged from alternate eyes on alternate presentations of the set of stimuli. Onset, duration, and offset of each fixation were recorded on an event recorder.

Procedure. The stimuli were presented one at a time, each for 18 seconds. A half-second blackout of the screen served as the interstimulus interval. The stimuli were arranged in a different random order for each presentation of the set of stimuli. Presentations of the stimuli were continued without interruption until the infant cried or fell asleep or until each of the stimuli had been presented seven times. Data from a *S* were analyzed only if he viewed at least three presentations of each stimulus. Thus, each *S* received a minimum of three and a maximum of seven trials.

RESULTS

The two commonly used measures of fixation time, duration of the first fixation in each stimulus presentation, and total fixation time for each stimulus presentation, were taken from the record produced by the event recorder. Each of these measures was analyzed in two ways, with data from the first three trials alone, and with data from all trials which

the subject received. Since some of the subjects did not receive all seven trials, it was not possible to compare directly the first three trials with the last four trials.

The results for the four measures were similar, but some discrepancies did occur. The results will be presented first for the response measure which showed the greatest sensitivity to stimulus differences, i.e. the measure which produced the greatest number of significant differences with the planned comparisons. The discrepancies which occurred with the other three response measures will then be indicated.

Because the fixed stimulus presentation time imposed a floor (0 seconds) and a ceiling (18 seconds) on the fixation times, there resulted skewed distributions for some stimuli and for some *S*s. Medians, which are more representative than means of the central tendency of skewed distributions, were taken for each stimulus for each *S*. The distributions of these medians were more reasonable approximations of the normal distribution. Age-group means of these medians were computed for each stimulus, and conventional parametric statistics were performed. The mean median fixation time for each of the four response measures is presented in Table 2.

Total fixation time across all trials. The mean median fixation time over all trials is presented in Table 2. An analysis of variance of these data showed the main effects of age ($p < .05$) and stimuli ($p < .01$) and the age by stimuli interaction ($p < .05$) to be significant. The results of the planned comparisons are as follows.

Schematic and Scrambled were not significantly different from each other at any age level. Neither was there a significant difference between Eyes and Mouth-Nose at any age. For all three age groups both complexity comparisons (Comparisons V and VI) were significant, indicating a preference for the more complex stimuli. However, it should be noted in Table 2 that the fixation time for Blank was approximately equal to that for Eyes and Mouth-Nose in the 4-week-old group. Developmental changes are evident in the relationships among the three complete, unaltered faces. In the 4-week-old group, Drawing produced a lower score than did Schematic and Photo (Comparison IV, $p < .01$). In the 10-week-old group there were no significant differences among these stimuli. In the 16-week-old group Photo was preferred to the other complete faces (Comparison III, $p < .05$).

Total fixation time across the first three trials. The only major difference between this response measure and total fixation time over all trials occurred in the youngest group. The two comparisons along the complexity continuum did not produce significant results for the 4-week-old group. This response measure produced only one significant difference between stimuli in this group, Scrambled, Schematic and Photo being

TABLE 2

RESULTS: WITH EACH OF THE FOUR RESPONSE MEASURES GIVEN IN MEAN MEDIAN FIXATION TIME IN SECONDS

	Blank	Mouth-Nose	Eyes	Scrambled	Schematic	Drawing	Photo
4-Week-old group							
Total, all trials	4.76	4.92	6.12	8.75	8.60	5.30	9.30
Total, three trails	6.85	5.25	7.30	7.60	9.80	6.10	10.20
First fixation, all trials	2.70	2.27	2.52	4.95	5.00	2.40	3.62
First fixation, three trials	3.40	2.70	4.00	4.60	5.45	3.95	5.05
10-Week-old group							
Total, all trials	3.05	8.10	8.07	12.97	11.57	11.07	13.62
Total, three trials	4.10	8.85	9.60	12.45	11.75	10.95	12.35
First fixation, all trials	1.20	4.07	4.80	10.05	8.80	6.07	9.22
First fixation, three trials	1.10	4.65	4.85	10.25	8.95	4.80	8.45
16-Week-old group							
Total, all trials	5.55	9.82	10.17	11.22	12.62	12.15	14.57
Total, three trials	5.25	10.00	11.80	13.00	12.75	14.40	15.00
First fixation, all trials	1.72	3.90	5.22	4.67	7.00	6.60	9.32
First fixation, three trials	1.80	5.30	5.30	4.65	8.80	9.50	8.55

preferred to Drawing (Comparison IV). The results for the two older groups were the same with this measure of fixation time as they were with total fixation time across all trials.

Duration of first fixation across all trials. The remaining two measures of fixation time were based on the duration of the first fixation which occurred in each stimulus presentation. It should be noted that a fixed stimulus presentation time was used in this study, as was done in previous studies which compared response measures (Lewis, Kagan, and Kalafat, 1966). Therefore, the first fixation in a stimulus presentation was occasionally terminated by the removal of the stimulus rather than by the infant himself. This truncation of the first fixation occurred on 12% of the stimulus presentations for the 4-week-old group, on 22% for the 10-week-old group, and on 17% for the 16-week-old group.

Duration of the first fixation over all trials showed no significant stimulus effects for the 4-week-old group. The results for the 10-week-old group were somewhat discrepant from the results using total fixation time. The two complexity comparisons were significant, indicating a preference for more complex stimuli as was found before. There was also a significant difference between Drawing and the other complete faces, an effect which was found in the 4-week-old group with total fixation time but not in the 10-week-old group. For the 16-week-old group, the planned comparisons produced the same results as were found with total fixation time. However, there was essentially no difference between the means for Eyes and Mouth-Nose and the means for Scrambled and Schematic. Therefore, it is concluded that the complexity effect was confined to a preference for the faces containing some or all features over the blank oval outline.

Duration of first fixation across the first three trials. The results with this measure of fixation time are the same as those for duration of first fixation with data included from all trials with one rather striking exception. In the 16-week-old group the means for Schematic and Drawing were higher and the mean for Photo was lower than the values obtained when data from all trials were included. This resulted in a significant difference being obtained between Scrambled and Schematic ($p < .05$) and no significant difference being obtained between Photo and the mean of Scrambled and Schematic. The results for the other two groups are essentially the same as were obtained with duration of first fixation over all trials.

DISCUSSION

Given the ambiguities caused by the differences between measures of fixation time, it is still possible to draw conclusions about the visual

preferences shown in these data. Complexity, defined as the number of facial features contained within an oval outline, proved to have an effect at all three age levels, with all three groups preferring the more complex to the less complex stimuli.

Comparisons among the three complete, unaltered faces showed developmental changes in visual preferences. The oldest group showed a preference for Photo over the two drawings with three of the four response measures. This result suggests a preference for more realistic representations of the face at 16 weeks of age. However, the younger subjects showed a preference for both Photo and Schematic over Drawing (at 4 weeks with total fixation time, and at 10 weeks with duration of first fixation). This preference ordering does not follow the continuum of realism which these three stimuli were intended to represent and probably should be attributed to some factor other than faceness which varied among these stimuli. It may have resulted from the fact that the realistic drawing contained more gray and less black and white and therefore less area of sharp contrast between black and white than did the schematic drawing and the photograph.

No evidence was found that infants 4 or 10 weeks of age prefer a facelike pattern to a non-facelike pattern equal in complexity. The data for the 16-week-old group are inconclusive on this point, but since the preference for Schematic over Scrambled showed up ($p < .05$) in only one of the four response measures and occurred in a response measure which was not entirely appropriate to the procedures used, it cannot be said that the preference, if it does exist, is a very strong one. Koopman and Ames (1967) also report finding no significant difference between the schematic face and a symmetrically scrambled control stimulus in a study conducted at the same time as this one. These independent replications of this result add weight to the conclusion that infants have little or no preference for a facelike pattern per se.

No evidence was found with any response measure that infants at any of the three ages look at eyes more than at the features in the lower half of the face or vice versa. Since Ahrens (1954) reports that infants up to three months old smile more at a face with no mouth or nose than at a face with no eyes, these data suggest that smiling and visual preferences do not follow the same developmental sequence and may not be controlled by the same processes.

The results show that the visual preferences of these infants among these stimuli were not strongly influenced by the degree of faceness of the various stimuli. Only questionable evidence was found of a preference for Schematic over Scrambled, and at only one age level. The preferences which were found among the three faces varying in degree

of realism were dependent upon the response measure analyzed and did not consistently follow the realism continuum. Complexity seems to have dominated as the primary determiner of visual preferences in this set of stimuli. Preferences for the stimuli with more elements was consistently found at every age level and with every response measure at the two older ages. The visual preferences found in this study seem to have been controlled more strongly by complexity and perhaps contrast than be degree of resemblance to a human face. This finding recommends caution in the interpretation of preferences found among representations of faces. Adequate control of other variables such as complexity and contrast should be attempted, or the extent to which they influence the results should be assessed, before the results are attributed to a response to the face as a meaningful stimulus.

It is evident that conclusions about the visual preferences shown in these data depend somewhat on the response measure used. In the 4-week-old group, significant differences between stimuli were found only with total fixation time, and not with duration of first fixtion. In the 16-week-old group, conclusions about the preference for Schematic over Scrambled depend upon the response measure. As we have seen, however, when the data for all ages are considered, it seems not to be a simple case of one measure being uniformly most sensitive to stimulus differences. At both 4 and 10 weeks of age Drawing was looked at less than the other complete faces. This difference reached significance with a different response measure at each of the two age levels, with total fixation at 4 weeks and with duration of first fixation at 10 weeks.

This instance of definite differences among response measures would indicate that reports of studies of visual preferences should consider carefully the measures used. This is especially important for comparisons among studies, in which different experimenters, using different subjects and procedures, might be expected to produce even greater discrepancies.

REFERENCES

AHRENS, R. Beitrag zur Entwicklung des Physiognomie- und Mimikerkennens. *Zeitschrift für Experimentelle und Angewandte Psychologie,* 1954, **2**, 412–454. Translated under the title of 'Contribution on the development of physiognomy and mimicry recognition' by the National Institutes of Health Translating Service.

BERLYNE, D. E. The influence of the albedo and complexity of stimuli on visual fixation in the human infant. *British Journal of Psychology,* 1958, **49**, 315–318.

BRENNAN, W. M., AMES, E. W., AND MOORE, R. W. Age differences in infants' attention to patterns of different complexities. *Science,* 1966, **151**, 354–356.

FANTZ, R. L. The origin of form perception. *Scientific American,* 1961, **204**, 66–72.

FANTZ, R. L. Pattern discrimination and selective attention as determinants of per-

ceptual development from birth. In Kidd and Rivoire (Eds.), *Perceptual Development in Children*, New York: International Universities Press, 1966.

HAYNES, H., WHITE, B., AND HELD, R. Visual accommodation in human infants. *Science*, 1965, **148**, 528–530.

HAYS, W. L. *Statistics for Psychologists*. New York: Holt, Rinehart, and Winston, 1963.

HERSHENSON, M., MUNSINGER, H., AND KESSEN, W. Preferences for shapes of intermediate variability in the human newborn. *Science*, 1965, **147**, 630–631.

KAGAN, J., AND LEWIS, M. Studies of attention in the human infant. *Merrill-Palmer Quarterly*, 1965, **11**, 95–127.

KAGAN, J., HENKER, B. A., HEN-TOV, A., LEVINE, J., AND LEWIS, M. Infants' differential reactions to familiar and distorted faces. *Child Development*, 1966, **37**, 519–532.

KOOPMAN, P., AND AMES, E. Infants' preference for facial arrangements: a failure to replicate. Paper presented at meetings of the Society for Research in Child Development, New York, March, 1967.

LEWIS, M. KAGAN, J., AND KALAFAT, J. Patterns of fixation in the young infant. *Child Development*, 1966, **37**, 331–341.

LEWIS, M., CAMPBELL, H., BARTELS, B., AND FADEL, D. Infants' responses to facial stimuli during the first year of life. Unpublished manuscript, 1966.

THOMAS, H. Visual-fixation responses of infants to stimuli of varying complexity. *Child Development*, 1965, **36**, 629–638.

WATSON, J. S. Perception of object orientation in infants. *Merrill-Palmer Quarterly*, 1966, **12**, 73–94.

GOVT. OF W. BENGAL
LIBRARY
CALCUTTA-19

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 21–30 (1969)

Learning Set and Shift Behavior in Children[1]

ANNE SARAVO AND MAY KOLODNY
The Fels Research Institute

Twenty-five preschool children learned a series of dimensional shifts, four reversals (RS) and four nonreversals (NRS) followed by a conditional discrimination (CD) comparison of the two shifts. Evidence of learning set was obtained, and performance improved reliably on RS on both color and form dimensions. Facilitation on RS did not transfer to CD.

Both reversal shifts and learning set have been hypothesized to reflect differences in the ability to mediate (Reese, 1962; Kendler and Kendler, 1962; Tighe and Tighe, 1966); and the ease with which both tasks are accomplished seems to be related to both ontogenetic and phylogenetic differences (Miles and Meyer, 1956; Kaufman and Peterson, 1965; Kendler and Kendler, 1962; Gollin, 1964; Reese, 1963). There is disagreement, however, about the extent to which the mediation involved is specific to a particular developmental stage (Mackintosh, 1965; Kendler and Kendler, 1966).

Luria (1961) has summarized Russian research showing that deficiency in mediation can sometimes be overcome by special training; it is probable that, if the formation of a learning set facilitates mediation, or vice versa, the effects would be reflected in the relationship of the two shift paradigms. In addition, Mackintosh and Mackintosh (1964) suggested that improvement in habit reversal can be accomplished only at the cost of increased difficulty in shifting dimensions. This study was therefore designed to determine (a) the relationship between the ability to form a learning set using minimum stimuli (Riopelle, 1955) and the relative ease of reversal and nonreversal shifts, and (b) the changes with practice in the difficulty of dimensional shifting as compared with the difficulty of reversal.

METHOD

Subjects. The *S*s were 37 experimentally naive children attending nursery school in a day care center for culturally deprived, lower-class chil-

[1]Preparation of this report was supported by Grant HD-01570 to E. S. Gollin and by USPHS Post-doctoral Fellowship Grant 1-F2-MH-29, 557-01 to the senior author. Appreciation is due Mrs. Faye DeHaven, Director, Greater Dayton Child Development Center, for making *S*s available. The study was reported at the annual meeting, Eastern Psychological Association, Boston, Massachusetts, April, 1967.

a Psyl. Research

dren. Their ages ranged from 36–71 months; within this range there were 3-year-old (15), 4-year-old (12), and 5-year-old (10) children. The available Stanford-Binet IQ's ranged from 86–131, with a median of 103.

Apparatus and stimuli. A modified version of the Wisconsin General Test Apparatus was used. A black curtain with brightly colored nursery figures was interposed between *S* and *E* while the stimuli were baited; when the curtain was raised, *E*'s face was not visible to *S*. The stimulus objects were presented to *S* on a gray slide-out tray; the objects could be opened to reveal the candy rewards. The stimuli were wooden three-

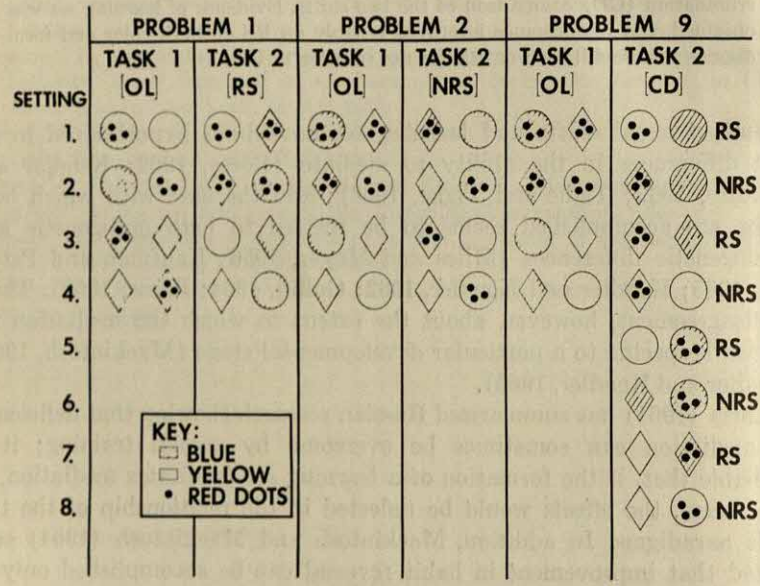


FIG. 1. Stimulus settings for one subject of problems 1, 2, and 9, illustrating original learning tasks (OL), reversal and nonreversal shifts (RS and NRS, respectively), and conditional discrimination (CD).

dimensional objects differing in form (circle or diamond), color (blue or yellow), and the presence or absence of three red dots.

Design. Each *S* learned a series of eight two-task transfer problems; half of these problems were reversal shifts, half were nonreversal shifts. This series was followed by a conditional discrimination problem (CD) which constituted a mixed list, within-subject comparison of both RS and NRS paradigms (see Saravo, 1967). The CD task required a reversal on half of the trials and a nonreversal on the others. Figure 1 illustrates a specimen of the settings on problems 1, 2, and 9 for one subject; for purposes of illustration, the left stimulus of each pair is positive.

The eight dimensional-shift problems were arranged so that each of

the four possible combinations of RS, NRS, and color or form relevant appeared twice in the eight problems, and so that the first two problems for each *S* included both a reversal and a nonreversal shift. The *S*s within each age group were randomly assigned to one of eight counterbalanced sequences of these combinations. For example, the *S* whose settings appear in the figure received the following sequence of positive stimuli in successive problems: problem 1, blue (OL)—yellow (RS); 2, blue (OL)—diamond (NRS); 3, circle (OL)—yellow (NRS); 4, diamond (OL)—circle (RS); 5, diamond (OL)—blue (NRS); 6, circle (OL)—diamond (RS); 7, yellow (OL)—circle (NRS); and 8, yellow (OL)—blue (RS). In the ninth problem, blue was correct in the first task (OL), and both yellow (RS) and diamond (NRS) were correct in the CD task, as illustrated.

Each task involved all three dimensions; one relevant, one irrelevant, and the third varying successively; i.e., constant within a single trial setting. For the example in the figure in task 1, problem 1, color is relevant, dots are irrelevant, and form varies successively, that is, both stimuli are circles on one trial, and both are diamonds on another trial. Dots were never relevant. The effects of partial reinforcement and stimulus novelty were controlled by presenting the appropriate first task variable dimension as the successively variable (constant) dimension in the second task. In the nonreversal shift the previously relevant dimension was made constant; in the reversal shift, the previously irrelevant dimension was made constant. Stimulus sequences for each task were predetermined so that the spatial arrangements for each stimulus dimension followed a Gellermann order, and each value of a dimension was combined with every value on every other dimension an equal number of times.

Procedure. For the first eight problems and task 1 of problem 9, *S*s learned each task to a criterion of 9/10 correct. The *S*s were given a shift task immediately after reaching criterion on the first task of a problem; after *S*s reached criterion on a shift, the session was ended, and they began a new problem the next working day. Training on a particular task was continued for a maximum of 4 days, 36 trials per day. If *S* did not reach criterion within these 144 trials and/or a special training procedure described below, he was dropped from the experiment. On task 2 of problem 9 (CD), *S*s were run to a criterion of 9/10 correct or a maximum of 72 trials, whichever came first.

At the beginning of the series, the *S*s were seated in front of the apparatus and shown the four positive stimuli of the first task. The *E* explained, "These are my hiding places. They have surprises inside. Open them up and see what you find." After *S* consumed the candy, *E* presented the two stimuli of trial 1 and baited the positive stimulus in front

of *S*. The *E* then said, "Now only one has the surprise. Which one do you think it is?" After *S* picked the correct stimulus, *E* said, "That's right. Now I'm going to hide some more surprises and you see if you can find the hiding place every time. Remember to pick just one each time, the one with the surprise." The *S* was verbally reinforced for every correct response.

The special training procedure was given if *S* did not reach criterion on any task within four days. At the end of 36 trials on the fourth day, the *E* stacked the positive stimuli vertically on one side of the tray and the negative stimuli on the other side of the tray. Pointing to the positive stimuli, *E* said, "These are my hiding places, they have the surprises in them. These are not my hiding places, they do not have surprises in them. Open up my hiding places." After *S* saw the rewards, *E* baited the stimuli in front of *S* until *S* made four consecutive correct responses. The *E* continued with the baiting out of *S*'s sight for 36 more trials or until criterion was reached on the regular trials. Subjects were rewarded with a candy after each correct response, and a cartoon card at the end of the session every time criterion was reached on a particular task.

RESULTS

Task 1 Performance

Figure 2 presents the learning curve of the mean number of trials to criterion on original learning (OL) for all problems. One *S*'s unusually poor performance on problem 9 was excepted from the figure because it was neither typical of his past performance nor of the other *S*s; teacher's reports indicated an outside source of emotional disturbance on those days. His performance is not excluded from the statistics performed on the data, however.

Early in learning, repeated shifts hindered *S*'s over-all performance. A significant number of *S*s showed a decrement in OL performance in one or more problems when compared with the first problem (Binomial test, two-tailed, $x = 5$, $N = 25$, $p = .004$), even though the positive stimuli had been correct in at least one previous problem. For the 3-year-old *S*s, this negative transfer prevented most of them from finishing the series. Of the children who reached criterion on the first task, 19 of 22 older children but only 6 of the 15 three-year-olds could finish the series; this difference between age groups was statistically reliable ($\chi^2 = 6.17$, $df = 1$, $p < .02$). Only *S*s finishing the series were included in the over-all data analysis.

In general, the negative transfer produced early in learning was superseded by the formation of a learning set at the end of the series. Fewer trials were required to reach criterion on OL as the series progressed ($F = 8.41$, $df = 3/72$, $p < .001$). Evidence for learning set was also

demonstrated by a within-S comparison of OL performance on problems 1 and 9. It was found that a significant number of *S*s took fewer trials to reach criterion in the last problem (Binomial test, $x = 4$, $N = 25$, $p < .004$). There were no differences in the effect of successive problems as a function of relevant dimensions, nor were there any over-all differences between dimensions in ease of learning.

However, since after the first problem all OL tasks were actually shift tasks, only the first OL task provides an unambiguous measure of dimensional preference. Performance on the first OL also showed no reliable

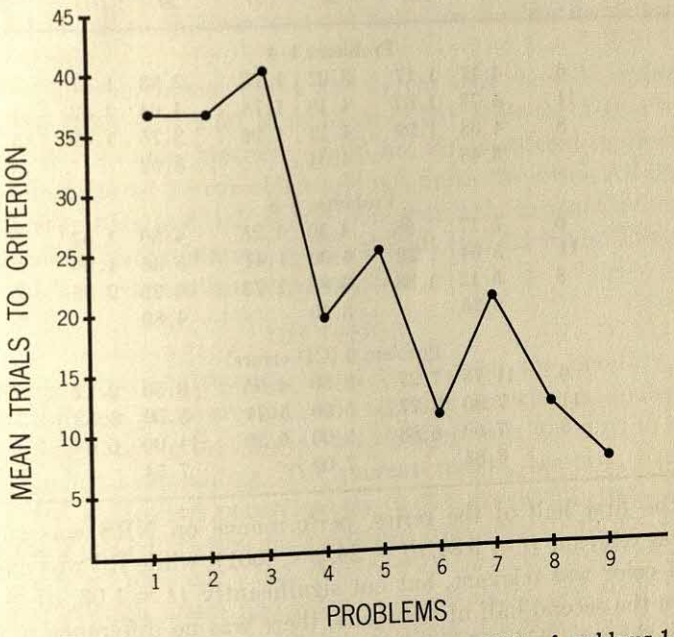


FIG. 2. Mean trials to criterion on the first tasks (OL) of problems 1–9.

differences as a function of relevant dimensions, although there were fewer trials to criterion and more correct responses on trials 2–8 inclusive if form was relevant.

Task 2 Performance

An analysis of variance was performed on the number of correct responses on trials 2–8 inclusive, with age, practice (problems 1–4 *vs.* problems 5–8), shifts, and dimensions as the variables in a $3 \times 2 \times 2 \times 2$ mixed design. This measure was used because the trials to criterion measure produced marked heterogeneity of variance in the second task. The direction of means, however, were congruent. Means and SD's of the number of correct responses appear in Table 1.

TABLE 1

MEANS AND SD'S OF THE NUMBER CORRECT ON TRIALS 2–8 INCLUSIVE FOR SECOND TASKS OF PROBLEMS 1–8, AND OF THE NUMBER OF ERRORS WITHIN 72 TRIALS ON CD

Age *Ss*, year	*N*	Dimension shifts: Color RS *M*	Color RS *SD*	Color NRS *M*	Color NRS *SD*	Form RS *M*	Form RS *SD*	Form NRS *M*	Form NRS *SD*
				Problems 1–4					
3	6	4.17	1.17	3.67	1.75	2.83	1.17	6.00	1.26
4	11	4.73	1.62	4.18	1.78	4.64	2.42	5.45	1.63
5	8	4.38	1.99	4.12	1.96	3.75	1.49	5.50	1.51
		4.48		4.04		3.92		5.60	
				Problems 5–8					
3	6	5.17	.98	4.50	1.25	4.50	1.52	5.00	2.00
4	11	5.64	1.29	6.00	1.41	5.36	1.80	5.82	1.33
5	8	5.12	1.88	5.88	1.73	4.25	2.25	5.12	1.36
		5.36		5.60		4.80		5.40	
				Problem 9 (CD errors)					
3	6	11.75	7.27	8.50	4.95	8.50	2.12	9.75	5.91
4	11	7.80	3.77	5.50	5.21	5.50	5.32	3.20	2.49
5	8	7.60	6.88	9.00	5.29	11.00	6.08	4.60	3.13
		8.86		7.00		7.54		5.57	

On the first half of the series, performance on NRS was superior if form was relevant ($t = 3.84$, $df = 24$, $p < .001$), while RS was easier than NRS if color was relevant, but not significantly ($t = 1.03$, $df = 24$, $p < .10$). On the second half of the series, there was no difference in performance on the two shifts. In addition, practice improved reversal performance when either form ($t = 2.17$, $df = 24$, $p < .05$) or color ($t = 2.37$, $df = 24$, $p < .05$) was relevant, while practice reliably improved dimensional shifting (NRS) only if the shift was to color ($t = 3.27$, $df = 24$, $p < .01$). The effect of practice then, was to make performance on the two shifts more alike with respect to the relevant dimensions. These effects are supported by a dimensions $\times$ shifts interaction ($F = 9.33$, $df = 1/154$, $p < .01$) and a dimensions $\times$ shifts $\times$ practice interaction ($F = 4.21$, $df = 1/154$, $p < .05$). Performance also improved reliably over-all ($F = 10.52$, $df = 1/154$, $p < .01$), and RS was more difficult than NRS over-all ($F = 5.91$, $df = 1/154$, $p < .05$). There were no over-all differences in task 2 performance as a function of relevant dimensions, however.

A finer analysis of the shift effects over problems was obtained by recombining the data into the sequence they were given. NRS superiority was significant only in the first two problems ($t = 2.11$, $df = 24$, $p < .05$). Thereafter, the children performed with equal facility on the two shifts. As learning on the series progressed, performance improved on RS from problem 1 to problem 8 (Binomial test, $x = 2$, $N = 12$, $p = .04$, N equals the number who had RS on problem 1) while NRS performance was not reliably affected. There were no differences in the original relationship of the two shifts as a function of ability to form a learning set, that is, children who showed facilitation on the series performed the same on the first shifts as children who did not show facilitation or could not finish the series.

On the CD comparison, the 4 and 5-year olds showed superior performance on NRS ($t = 2.50$, $df = 18$, $p < .05$) while the 3-year olds did not. These 3-year olds, however, could not be considered representative since the majority of 3-year olds could not finish the series. All but three *Ss*, two 3-year olds and one 5-year old, were able to reach criterion on the conditional discrimination within the allotted 72 trials. The median trial on which a criterion run began was 27.

DISCUSSION

The early negative transfer found in OL and the inability of the youngest children to overcome this negative transfer is analogous to Gollin's (1964) finding that 3-year-old children were hindered in reversal performance by overtraining. The present results, however, cannot be explained by a Hullian perseveration to the previously rewarded cue. Most of the 3-year olds who were dropped from the experiment failed on or after the third problem. Thus, they had already demonstrated successful solution of both intra- and extradimension shifts, and some had had experience with the positive stimuli. Without exception, they failed to meet criterion on the OL part of the problem, which suggests that delay between problems actually increased their confusion. Their performance within the last, noncriterion task showed perseveration to all of the varying dimensions in turn, sometimes with runs of 10 or more to the same cue. Although some of the nonlearners did exhibit emotional responses after having repeatedly picked the incorrect stimulus, most of them subsequently asked to continue the "games." This indicates that the reason the 3-year olds could not finish the series was not simply motivational. In addition, there was little suggestion that the type of shift from the previous task (task 2 on the preceding problem) had an effect on their performance. Five out of 9 were assigned the opposite cue on the same relevant dimension as the preceding task, corresponding to a reversal.

Reese (1962, p. 503) assumed that verbal mediators were involved in discrimination set, (which he distinguishes from learning set (Reese, 1963, p. 117)) and in the efficient performance on the shift paradigms. If this assumption is correct, then the mediational deficiency hypothesis would predict a relationship between developmental level, ability to form a discrimination set, and the relative ease of the two shifts at the beginning of the learning series. The ability to establish the discrimination set seemed to be related to age, supporting Reese, but this relationship is somewhat tenuous since its sole support rests on the inability of the 3-year olds to finish the series. Furthermore, there was apparently no relation between the relative ease of the two shifts at the beginning of the series and the ability to form a learning, or discrimination set. Thus, if the ability to mediate, as deduced by the establishment of a discrimination set, is a function of developmental level, it was not reflected in the original shift performance of the children in this study. If, however, both discrimination set and reversal facilitation are dependent on the development of observing responses, then according to attention theory (Zeaman and House, 1963), performance on reversals would be expected to improve concomitantly with OL performance, as was the case.

Although Mackintosh (1962) reported that ease of reversal learning was inversely related to ease of dimensional shifting, the present study failed to find any over-all effect on dimensional shifting, even though reversing within a dimension was facilitated. Furthermore, both RS and NRS performance were facilitated if color was the relevant dimension. Since improvement on OL was found in both dimensions, improvement in habit reversal cannot be attributed to the development of a particular dimensional set, or an increase in strength of one relevant stimulus analyzer. Indeed, the effect of practice was both to weaken any preexperimentally acquired dimensional preference and to facilitate reversals.

This interaction of shifts, dimensions, and practice seems in part paradoxical. If some of the children could reverse on one dimension in the beginning of learning, why did this initial advantage not become more pronounced?

Zeaman and House (1963, pp. 219–220) noted that learning set could result from "extinction of observing responses to the class of dimensions which are never relevant and the acquisition of strong tendencies to observe those dimensions which are frequently relevant".[2] As the series

[2] Interproblem improvement might also be explained, in very much the same terms, by a one-stage discrimination model (e.g., Spence, 1936) if it is assumed that as training progresses, values become equivalent on dimensions which are never relevant. Such nonrelevant dimensions, then, should cease to have functional control over discriminative responses. This kind of formulation might also be extended to include most "dimensional" responding (Saravo, 1965, p. 45).

progressed, and the third irrelevant dimension (dots) was never reinforced, *Ss* probably stopped attending to this dimension. By the end of learning, each task came to represent only a choice between the two cues of the relevant dimension, except in those tasks in which both color and form varied simultaneously. This, in part, probably contributed to the facilitation in OL over problems. Furthermore, if this analysis is correct, mediational behavior (in the form of attention to the relevant dimension) would produce no advantage in reversal shifting over dimensional shifting since only the relevant dimension would be attended to by most *Ss* in either RS or NRS. Moreover, the failure to transfer reversal facilitation to the CD task would then be due to the recovery of extinguished attentional responses to irrelevant cues. Thus, changing the type of problem acted as a "disinhibitor" in the same sense as novel stimuli.

Although attention theory can predict facilitation of reversal performance, it can not also explain the increased difficulty of the 3-year olds, nor the negative transfer produced by reversal shifts at the beginning of learning. This study and others (see Wolff, 1967) seem to indicate that, although dimensional responding does not seem to be restricted to one developmental level, an initial tendency to cue perseverate may be age related.

REFERENCES

Gollin, E. S. Reversal learning and conditional discrimination in children. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 441–445.

Kaufman, M. E., and Peterson, W. M. Acquisition of a learning set by normal and mentally retarded children. *American Journal of Mental Deficiency*, 1965, **69**, 865–870.

Kendler, H. H., and Kendler, T. S. Vertical and horizontal processes in problem solving. *Psychological Review*, 1962, **69**, 1–16.

Kendler, H. H., and Kendler, T. S. Selective attention versus mediation; Some comments on Mackintosh's analysis of two-stage models of discrimination learning. *Psychological Bulletin*, 1966, **66**, 282–288.

Luria, A. R. *The role of speech in the regulation of normal and abnormal behavior.* New York: Liveright, 1961.

Mackintosh, N. J. The effects of overtraining on a reversal and nonreversal shift. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 555–559.

Mackintosh, N. J., and Mackintosh, J. Performance of *Octopus* over a series of reversals of a simultaneous discrimination. *Animal Behavior*, 1964, **12**, 321–324.

Mackintosh, N. J. Selective attention in animal discrimination learning. *Psychological Bulletin*, 1965, **64**, 124–150.

Marsh, G. Effect of overtraining on reversal and nonreversal shifts in nursery school children. *Child Development*, 1964, **35**, 1367–1372.

Miles, R. C., and Meyer, D. R. Learning sets in marmosets. *Journal of Comparative and Physiological Psychology*, 1956, **49**, 217–222.

Reese, H. W. Verbal mediation as a function of age level. *Psychological Bulletin*, 1962, **59**, 502–509.

REESE, H. W. Discrimination learning set in children. In L. P. Lipsitt and C. C. Spiker (Eds.), *Advances in Child Development and Behavior.* Vol. I. New York: Academic Press, 1963, p. 117.

RIOPELLE, A. J. Learning sets from minimum stimuli. *Journal of Experimental Psychology,* 1955, **49,** 28–32.

SARAVO, A. Effect of number of variable dimensions on reversal and nonreversal shifts. *Journal of Comparative and Physiological Psychology,* 1967, **64,** 93–97.

SARAVO, A. Effect of number of alternative dimensions on the learning of reversal and non-reversal shifts in a within subjects design. (Doctoral dissertation, University of Massachusetts) Ann Arbor: University Microfilms, 1965.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychological Review,* 1936, **43,** 427–449.

TIGHE, T. J., AND TIGHE, L. S. Overtraining and optional shift behavior in rats and children. *Journal of Comparative Physiological Psychology,* 1966, **62,** 49–54.

WOLFF, J. L. Concept-shift and discrimination-reversal learning in humans. *Psychological Bulletin,* 1967, **68,** 369–408.

ZEAMAN, D., AND HOUSE, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency.* New York: McGraw-Hill, 1963, pp. 159–223.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 31–35 (1969)

Age, Cognitive Control and Extinction[1]

ROY GLADSTONE
Oklahoma State University

Two groups, one composed of 2½–3½ year old children, the other of 4½ to 5½-year-old children, were given a simple operant task and were rewarded on a variable ratio five schedule. The rewards and their eventual exhaustion were visible. The younger *Ss* gave significantly more extinction Rs than the older *Ss*. Several explanations of the results are presented. The explanations revolve about the variables of ongoing behavior, reinforcement schedule, discrimination, total display versus critical cue, integration across space, *E* control, and associationistic versus cognitive constructs.

This experiment was designed to discover the approximate age at which nonresponse extinction starts to occur. A pilot experiment (unpublished) suggested that a cue which is susceptible to a logical interpretation exercises little or no control over the behavior of children of about 3 years old although it exercises quite strong control over the behavior of children of 5–6 years. A more closely controlled experiment confirmed that finding and suggested that a point of inflection exists in the age range 3½–4½ years of age (Gladstone, 1966a). This experiment is a replication of the latter.

A discussion of pertinent theories will appear after a description of the experiment.

EXPERIMENT

Subjects

Two groups were used, one composed of 14 children from 2½ to 3½ years of age, the other composed of 21 children from 4½ to 5½ years of age, all Caucasian. It is likely that both socio-economic level and MA were above average in both groups although neither was measured.

Equipment

A rat pellet feeder was used with 10 BBs substituted for rat pellets. The BBs were all clearly visible to *S* through a piece of plate glass. The

[1]I am most grateful for aid which was given to me on this project. Miss Joyce Chien did practically all the work of securing subjects and running them. The work out of which this experiment evolved was supported by a U. S. Office of Education grant.

feeder was operated by a light switch, one on-off sequence composing one R. In each case 50 Rs fed out all 10 BBs.

Rewards

Rewards were pennies, gum, candy, and small toys.

Directions

E used the following directions, the portion in quotation marks being directed at *S*.

"Do you see all these prizes here? In this dish we have some pennies (pause) in this dish we have some little toys (pause) and in this dish we have some gum. Would you like to have some of these prizes?" (Indicate all the dishes). "All right, you can get some."

"I will give you your choice of one of these prizes when you get a BB out of this machine (point to BBs). Each time you get a BB I will let you take one of whatever you want. Now can you tell me how you get prizes?" If he knows, say "Good," if not, explain further.

"All right, here is how you get the BBs. Press this clicker here and pretty soon a BB will drop into this tray. Like this." (*E* uses the light switch to produce one BB.) "See—now you have one BB. Which prize would you like to have for your BB?" (Give him the prize he selects). "All right, now you do it. Tell me when you are finished pressing the clicker."

With each BB ask him to choose a prize. If he stops without saying anything say, "Would you like to go on or are you finished pressing the clicker?"

Two *S*s refused to stop to choose a prize after getting one BB even through *E* asked *S* what prize he would like. They were allowed to continue.

Theory

When the last BB falls from the feeder the visual display is changed slightly but critically. It was hypothesized that the older group would use that information to reduce their extinction Rs severely, many of them to zero, and that the younger group would not be able to use that information as well. It is suggested that extinction behavior may be a function of many influences and that the information contained in the visual display may interact with other influences to reduce rather than eliminate extinction Rs. If none of the children were to yield an extinction score of zero, the cognitive explanation would be weak at best.

The specific hypothesis is that the older group will yield fewer extinction Rs than the younger (a directional hypothesis).

RESULTS

The older children yielded statistically significantly fewer extinction Rs than the younger ($p = .023$, one-tailed test) by the Mann-Whitney

U Test (Siegel, 1956). Since the number of zero-level extinction Rs is important, all the data are given in Table 1.

TABLE 1
NUMBER OF EXTINCTION RS FOR EACH *S*[a]

2½–3½ years		4½–5½ years		
1	4	31	0	6
3	2	5	1	97
11	3	0	0	6
1	52	11	0	0
55	56	0	0	53
10	19	3	0	1
28	2	1	0	168

[a] While the BB fell when the switch went from "on" to "off" most *S*s flicked the switch back to "on" (half an R) on a significant number of trials. This is presumed to be a matter of the time necessary to make a decision and to inhibit an on-going, rapid act rather than a failure to make a decision. If *S* stopped at ½ an extinction R it was recorded as zero.

CONCLUSION

As expected, a large percentage of older *S*s gave zero extinction Rs. During the period when the child goes from 3½ to 4½ years of age something happens to the typical Caucasian (possibly fairly bright) child which makes his extinction behavior much more susceptible to control by the logical inference which may be derived from a rewards-all-gone display. This conclusion is considered by this investigator to be greatly strengthened by virtue of the fact that it has now been found three separate times by three different *E*s.

The implication is very strong that there is a rather sharp discontinuity which appears sometime in the age range 3½–4½ years during which time cognition achieves major control over the behavior of the typical *S* in the situation used.

DISCUSSION

These results may be explained in several ways. One is in terms of shift of control from perseveration to control by cognition.[2] Kendler and Kendler (1964) suggest that in younger children verbal responses and overt choice may be parallel processes which go on independently. Luria (1961) indicates that directions from an adult can start a behavior

[2] Reese (1962) has reviewed the literature on verbal mediation as a function of age level.

sequence in children but cannot stop it. The inference from both these sources is that the younger children of this experiment may have recognized the illogic of their extinction behavior but were unable to stop it. Indeed some *Ss* in this study acted as if this were the case saying there were no more BBs while giving Rs.

A second, related possibility is that the age at which cognition achieves control is a function of the particular schedule of reinforcement used.

A third possibility is that the younger *Ss* were unable to perceive the critical cue—the emptiness of the reward reservoir. The discomfort of some of the *Ss* in continuing (also remarked by Luria) militates against this theory.

A fourth possibility is that the younger *Ss* were controlled by the large mass of constant cues while the older were able to respond to the smaller cue of change from BB present to no BBs. This is an explanation in terms of a modification of the discrimination hypothesis (Bitterman, Fedderson, and Tyler, 1953).

A fifth theory with no evidence in its favor is that young *Ss* are especially handicapped in using information from a visual display effectively when the reward is removed in some dimension from the operant behavior. Pertinent dimensions are space, time, complexity and attention.

A sixth explanation is that the behavior of *S* was under the control of *E* and *E* unconsciously pressured the children to respond as they did. The fact that the phenomenon appeared under three separate *Es* militates against this theory.

Finally, while associationistic constructs can be used to explain the results (*S* learns not to respond to the cue of emptiness) this experiment and others (Gladstone 1966a, 1966b, Gladstone and Miller, 1968) were suggested by the cognitive viewpoint.

REFERENCES

BITTERMAN, M. E., FEDDERSON, W. E., AND TYLER, D. W. Secondary reinforcement and the discrimination hypothesis. *American Journal of Psychology*, 1953, **66**, 456–464.

GLADSTONE, R. Effect of extrinsic rewards on motivation. Cooperative Research Project, #5-8164-2-12-1, U. S. Office of Education, 1966a.

GLADSTONE, R. The significance of the extinction situation. *Journal of Psychology*, 1966b, **64**, 23–28.

GLADSTONE, R., AND MILLER, M. Logic, motivation, and extinction. *Journal of Psychology*, 1968, **68**, 33–38.

KENDLER, H. H., AND KENDLER, TRACY S. Vertical and horizontal processes in problem solving. *Psychological Review*, 1962, **69**, 1–16.

LURIA, A. R. The development of the regulatory role of speech. In R. J. C. Harper, *et al.* (Eds.). *The cognitive processes.* Englewood Cliffs, N. J.: Prentice Hall, 1964. Pp. 601–622.

REESE, H. W. Verbal mediation as a function of age level. *Psychological Bulletin,* 1962, **59**, 502–509.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 36–51 (1969)

Operant Control of Infant Vocal and Motor Behavior

W. C. SHEPPARD
University of Oregon

Differential responding under discriminative control in two concurrent responses, vocalization and leg kick, was achieved with an infant less than 3 months of age.

Both vocalization and leg kick were separately reinforced, placed under schedule control, and integrated into a disjunctive discrimination schedule in which each response was reinforced in the presence of a different discriminative stimulus; reinforcement and extinction were reciprocal operations here since the S^D for one response was an S^Δ for the other response.

The acquisition of differential responding under discriminative control clearly established that the behaviors under study were operants.

The experimental task in studying infant operant behavior is threefold: first, one must select a response class that is not tied too closely to eliciting stimuli; second, it is necessary to identify a stimulus that will affect the probability of occurrence of the operant without also having strong eliciting properties for other perhaps incompatible responses; and third, it is necessary to demonstrate that the behavior change in the experiment is a function of operant as opposed to respondent control.

The requirement that the response selected for study not be closely tied to eliciting stimuli severely limits one's choice, since an infant has only a limited behavioral repertoire, much of which is reflexive or is subject to multiple sources of control.

The difficulty of identifying a reinforcing stimulus is compounded by the constraints imposed upon investigations of infant behavior. Primary reinforcers such as food, water, warmth, etc., that have been shown to be effective reinforcers for more mature humans and numerous other species require prior deprivation which many investigators are unwilling to use with human infants. The procedural difficulties involved in the delivery of the reinforcers has also been a limiting factor. A further constraint is that the reinforcing stimulus should not have strong eliciting properties for the behavior under study or other incompatible behaviors.

Finally, the demonstration of operant control requires that the possibility of alternative forms of control be rejected. As mentioned earlier,

the behavioral repertoire of the human infant is highly reflexive. It can be argued that in an operant conditioning experiment, in which stimuli are presented following a response, that the stimuli presented are in fact eliciting the behavior and that any increase in the frequency of responding is attributable to this.

Several previous studies of infant behavior (Brackbill, 1958; Rheingold, Gewirtz, and Ross, 1959) have been criticized for not separating the eliciting from the reinforcing function of the stimuli presented.

Recent research (Weisberg, 1963; Siqueland and Lipsitt, 1966) has been directed towards the refinement of experimental procedures and designs that avoid this problem.

The most convincing way of meeting this argument is the establishment of differential responding under discriminative control in two or more conditions in which the reinforcing stimulus is presented with equal frequency and only the contingencies of reinforcement are varied. Since the frequency of reinforcement is the same in all conditions any differential responding that occurs can be attributed solely to the control exercised by the contingencies of stimulus presentation.

This study was directed at utilizing this design to provide a demonstration of operant control over vocal and motor behavior.

METHOD

Subject

The second-born child of the *E* was the subject in this experiment. The delivery was full-term, normal and easy. The birth weight was 9¾ pounds, which is slightly above average for male infants in the Middlewest. He was 10 days old and weighed 9¾ pounds when the experiment began and was 3 months old and weighed 17 pounds when the experiment ended. According to a periodic series of medical examinations made by the University of Michigan Child Care Center, the infant grew and developed within the average range. He was breast-fed, had no digestive disturbances and experienced no illnesses for the duration of the study. The *S* slept and was cared for in a plexiglass "air-crib" (T.M.I.) that also served as the experimental chamber. The infant was frequently out of the crib on excursions with the family and for play experiences at home. When the infant was in the crib, bells, stuffed animals and mobiles were provided. The infant rarely cried and presented every evidence of being an unusually alert, comfortable, and happy infant.

Apparatus

The experimental chamber was a commercially available plexiglass "air-crib" (T.M.I.) that also served as living quarters for the infant.

The experimental chamber was a commercially available plexiglass external sounds.

During experimental sessions a 24 by 36-in. piece of red translucent plastic was placed along one side of the crib and a 100-watt red light was placed behind the red plastic.

Two response transducing devices were used. To measure the rate of vocalization a directional microphone (Altec, 633A) was placed in the crib approximately 4-in. from the infant's head. The microphone was connected to a magnetic tape recorder (Tandberg, Model 64) and also to a fast-acting voice-operated switch (Miratel, Model V-1005-X) which was sensitive to and could separate all responses with an inter-response time of 175 msec or greater.

The second response transducer was a micro-switch mounted behind a hinged panel, 8 in. high by 19 in. wide, that was placed at one end of the crib with the infant positioned such that he could kick it. The switch registered a response whenever it was depressed and released.

The outputs of both manipulanda were connected to control circuitry consisting of a cumulative response recorder (Gerbrands), timers, counters, and relay circuitry. The cumulative recorder stepped upward with each response, blipped down with each reinforcement, and marked condition changes. Counters recorded the total number of responses and reinforcements during each session. The recording and control circuitry were located in an adjoining room.

Four different potentially reinforcing events were used. A small, 1 × 1 × ¼-in., vibrator, electrically pulsed six times a second was taped to the infant's right palm. A thin, flexible cord was used, allowing freedom of movement. A second device consisted of an aluminum box, 6 × 4 × 2-in., with two small light bulbs, one red and one yellow, mounted on it. The bulbs flashed in alternation once every second. The box was placed at eye level beside the infant. The third device was a magnetic tape repeater (Mag-Matic) placed at the end of the crib. The tape used was a recording of the infant's mother talking to the infant while nursing. The mother continued to repeat the same monologue to the infant on subsequent feedings for the duration of the study.

A recording of the mother's heart-beat, the fourth device, was used once.

Stimulus and response events were programmed so that a response could be followed by no consequence, by 5 seconds of vibration, or by 5 seconds of flashing lights and the recording of the mother's voice. The red light behind the red plastic could be turned on or off during either reinforcement condition as a discriminative stimulus (S^D). The control equipment determined the sequencing of the experimental conditions.

Procedure

Before each experimental session the infant was briefly removed from the crib along with the bells, stuffed animals and mobiles, the experimental equipment was placed in the crib, and the infant was then returned to the crib. The infant was placed in a supine position for all sessions, clothed in a diaper and shirt, with the temperature regulated at 81°F.

Sessions were usually run daily although there were several longer lapses between sessions as well as two sessions occurring in a single day. Early mornings after feeding proved to be the best time for the sessions.

Since the vocal response of an infant can be elicited by a number of conditions, which would interfere with and confound parts of this study, it was extremely important to remove all of the antecedents of respondent crying; special attention to feeding, burping, amount of sleep, and schedule of defecation and urination was required with the timing of the experimental sessions determined by these factors. The infant was observed during sessions from a concealed vantage point. Wet diapers were changed immediately. If the infant began to defecate or fall asleep the experimental session was terminated and the data not counted. These events ocurred 2 and 3 times, respectively.

There were 67 full sessions. The duration of the first thirty sessions was 30 minutes each. Later sessions were 20 minutes each. Experimental conditions are described below in chronological order.

Vocalization

Operant level. The procedure described above was followed. The operant level of vocalizing was measured for five sessions during which the experimental apparatus was present in the crib. During these sessions no consequence followed the response.

Continuous reinforcement. Continuous reinforcement (CRF) contingent on vocalization was initiated during sessions 6 through 11. The red light (S^D) was on during all sessions.

The first reinforcement session explored the use of a recording of the mother's heart-beat as a reinforcer.

In sessions 7 through 11 the vibrator was presented for a duration of 5 seconds contingent on each response.

Intermittent reinforcement. The conditioning of vocalization initiated using CRF and reinforced by the vibrator for 5 seconds was extended during sessions 12 through 18 to intermittent reinforcement on a fixed-ratio of one reinforcement for every three responses (FR 3). The light was on during all sessions.

Multiple schedule [MULT FR 3 DRO]. The two components of the

multiple schedule used in sessions 19 through 33 were FR 3 and differential reinforcement of zero rate (DRO). The red light was correlated with the FR 3 component. The infant was reinforced with the vibrator for 5 seconds, while the red light was on, contingent upon the emission of three vocalizations. This contingency was in effect for a total of three reinforcements. The light then switched off and the infant was reinforced with the vibrator for 5 seconds if he remained quiet for 30 seconds (this was reduced to 15 seconds after session 20 and further reduced to 7 seconds after session 27). Every vocalization while the red light was off reset the timer. At the end of the silent period the infant was reinforced and the red light switched on indicating a return to the FR component.

Starting with session 28 the flashing lights and the recording of the mother's voice were used as reinforcers replacing the vibrator.

The flashing lights and the recording of the mother's voice were used as reinforcers for the duration of the study.

Multiple schedule [*MULT FR 3 NONCONT REIN*]. The two components of the multiple schedule used in sessions 34 and 35 were FR 3 and noncontingent reinforcement. The red light was correlated with the FR component and no light was correlated with the noncontingent reinforcement component. The infant was reinforced with both the flashing lights and the recording of the mother's voice for 5 seconds, while the red light was on, contingent upon the emission of three vocalizations for a total of three reinforcements. The red light then switched off and the infant was reinforced with the flashing lights and the recording of the mother's voice for 5 seconds every 10 seconds for three reinforcements independent of the infant's behavior. The red light then switched on indicating a return to the FR component.

Leg Kick

Operant level. The operant level of leg kicking was measured during sessions 36 and 37. During these sessions no consequence followed the response.

Continuous reinforcement. During sessions 38 and 39 CRF contingent on leg kicking was initiated. The red light was on during all sessions. Reinforcement consisted of 5 seconds of both the flashing lights and the recording of the mother's voice.

Intermittent reinforcement. The conditioning of leg kicking under CRF was extended during sessions 40 through 43 to FR 3 and then to FR 5 during session 44 and 45. The red light was on during all sessions.

Extinction. During sessions 46 and 47 no consequence followed the leg-kick response. The red light was off.

Reconditioning. During session 48 leg kicking was reinforced on an intermittent schedule of FR 5. The red light was on.

Continuous reinforcement. CRF contingent on vocalization was reintroduced during session 49. The red light was off.

Intermittent reinforcement. Vocalization was reinforced on a FR 3 schedule during session 50 and 51. The red light was off.

Vocalization and Leg-Kick

During sessions 52 through 67 both vocalization and leg kick were reinforced on two concurrent multiple schedules.

During sessions 52 through 54 in the first 10 minutes of the session vocalization was reinforced on a FR 3 and the infant was moved away from the leg-kick apparatus which was inactive. In the second 10 minutes of the session the microphone was removed and inactive, and the infant was moved near the leg-kick apparatus and leg kicking was reinforced on a FR 3. The red light was correlated with the leg-kick component.

In sessions 55 through 58 both responses were reinforced as above but the reinforcement conditions were alternated every 5 minutes. Both the presence of the microphone and the proximity to the leg-kick apparatus during the periods when they were inactive were faded in by allowing the infant to be closer to them during each successive session.

In sessions 59 through 67 both manipulanda were present. Responses were reinforced as above with the reinforcement conditions alternating every 5 minutes. During the first and third 5-minute periods vocalization was reinforced on a FR 3 and leg kicking was extinguished. In the second and fourth 5-minute periods leg kicking was reinforced on a FR 3 and vocalization was extinguished. The red light was correlated with the leg-kick component. Reinforcement and extinction are reciprocal operations here since the stimulus correlated with reinforcement (S^D) for one response is correlated with extinction (S^Δ) for the other response, and vice-versa. Both the number of reinforced responses (S^D responses) and the number of unreinforced (S^Δ responses) during all sessions were recorded.

RESULTS

Vocalization

Figure 1 shows the cumulative response records of vocalization in sessions 1 through 5, operant level; sessions 7 through 11, continuous reinforcement (CRF); and sessions 12 through 18, fixed-ratio reinforcement (FR 3). The average number of responses during the operant

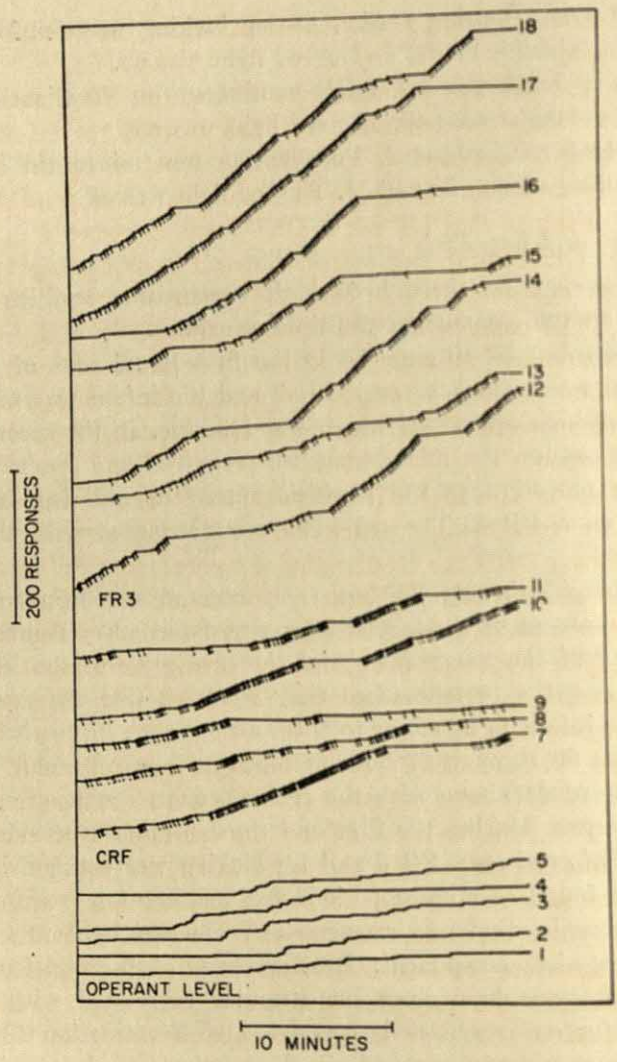


FIG. 1. Cumulative response records of vocalization in sessions 1 through 5 (operant level), 7 through 11 (CRF), and 12 through 18 (FR 3). Age (10-23 days).

level sessions was 43, during CRF sessions it was 100, and during FR 3 sessions it reached 215. Responding in both CRF and FR 3 sessions is marked by bursts of rapid responding followed by pauses. Towards the end of many sessions the response rate declines. This is perhaps attributable to satiation or possibly fatigue. The most common responses in these sessions were voiced and of short duration, although a number of other sounds such as coughs, snorts, and sneezes also occurred. No attempt was made to separate these. Cries, whines, and whimpers rarely

occurred. Little responding was observed while the reinforcement was being delivered.

Although the average response rate during the FR 3 sessions is five times greater than the operant level, this cannot yet be attributed to the reinforcing as opposed to the eliciting function of the stimuli presented. To remove this ambiguity by separating these possible functions a multiple schedule (MULT FR 3 DRO) was introduced in sessions 19 through 33. Differential responding during the two components of the schedule was not achieved. In fact the response rate remained greater during the DRO component. The time requirement initially selected for the DRO component, 30 seconds, proved to be much too long. That is to say, the infant seldom satisfied the criteria for reinforcement during the DRO component. DRO, therefore, operationally resembled experimental extinction.

Vocalization in the first FR component of the schedule was similar to that in the conditioning phase. However, after several unreinforced responses during DRO the topography of the response changed to crying. Attempts to remove this problem by shortening the DRO requirement from 30 seconds to 15 seconds and later to 7 seconds failed to eliminate the crying. The apparent elicitation of crying resulting from unreinforced responding in DRO completely interfered with the establishment of discriminative behavior.

In sessions 34 and 35 a second multiple schedule (MULT FR 3 NONCONT REIN) was introduced. Differential responding was not established. The response rate during the noncontingent reinforcement component was approximately the same as during the FR component. The response rate during the noncontingent reinforcement phase did not decrease since the infant continued to respond as in the FR component and therefore received reinforcement following vocalization. This served to increase the response rate and crying soon developed. The development of crying precipitated the abandonment of this schedule after two sessions. The research strategy was shifted at this point to achieving control over two operants, leg kick and vocalization, and then integrating these together into a schedule in which both operants are under discriminative control.

Leg Kick

Figure 2 shows the cumulative response records of sessions 36 through 48 in which the leg-kick response was studied. The average number of responses during operant level (sessions 36 and 37) was 15. CRF contingent upon leg kick (sessions 38 and 39) increased the average number of responses per session to 78, intermittent reinforcement (sessions 40–

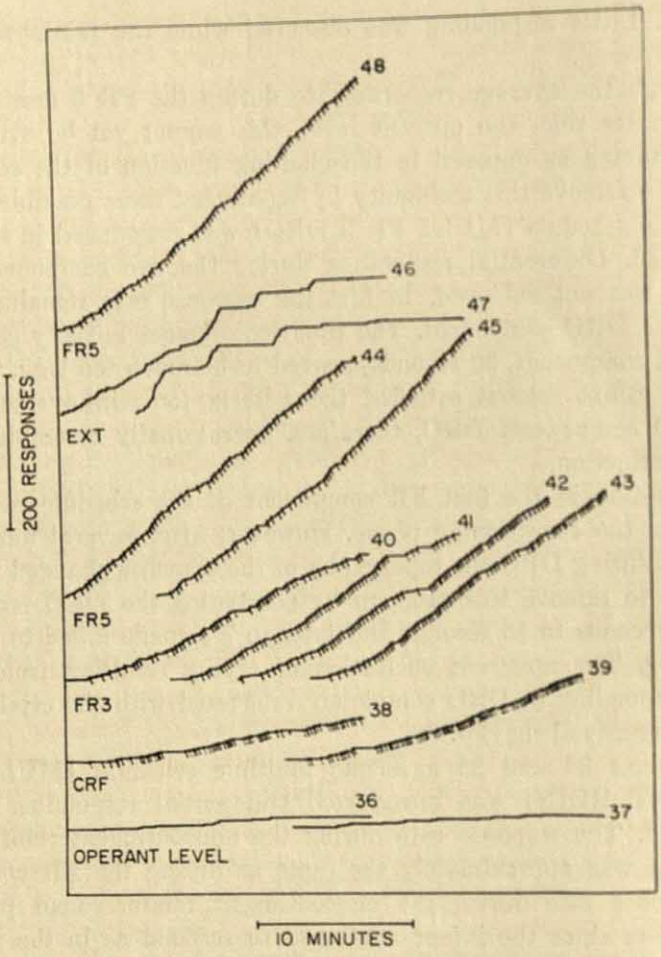


FIG. 2. Cumulative response records of leg kick in sessions 36 and 37 (operant level), 38 and 39 (CRF), 40 through 43 (FR 3), 44 and 45 (FR 5), 46 and 47 (EXT) and 48 (FR 5). Age (40–59 days).

45) further increased this to 193 during the FR 3 sessions and 310 during the FR 5 sessions. The average inter-response time[1] for the FR 5 sessions is approximately 3 seconds. This high response rate was maintained for the entire 20-minute sessions with little apparent fatigue. Only a few, short, postreinforcement pauses can be seen in Figure 2.

The topography of the response became highly differentiated. Two

[1] The average inter-response time was arrived at by subtracting the time the reinforcement was being presented (62 reinforcement $\times$ 5 sec = 5:10) from the duration of the experimental session (20 min — 5:10 min = 14:50 min) and dividing the average number of responses by this time (310/14:50 = 2.87 sec).

distinct response patterns developed. The first was alternation with one foot pressing and releasing the panel followed by the other foot pressing and releasing the panel. The responses were quite vigorous, to the point of producing loud noises and vibrating the crib. During early sessions the infant occasionally kept one leg extended against the panel and failed to release the micro-switch and thereby register a response. This latter dropped out and the response became a smooth extension, flexion movement. The second response pattern was executed with one foot, typically the right, resting on the heel with the toes in contact with the panel. A slight depression of the panel and its release was accomplished with little effort by the toes alone. Prolonged depression of the panel was not observed to occur as often as with the vigorous kicks. The second response pattern appeared to be somewhat more common although no measure of their relative frequency was made. There was little responding during the time when the reinforcement was being presented. However, in latter sessions the infant would often depress the panel during reinforcement and release it as soon as the reinforcement terminated, thereby registering a response.

Extinction of the response can be seen in Figure 2. Examination of the cumulative response records of sessions 46 and 47 shows that most of the responses in extinction, an average of 142 responses per session, occurred early in the session in rapid bursts followed by prolonged breaks in responding. This is consistent with what is characteristically found in extinction following fixed-ratio reinforcement (Ferster and Skinner, 1957, p. 57). Spontaneous recovery at the beginning of the second extinction session is evident. During extinction the response topography shifted to vigorous kicking. The infant engaged in a number of alternative behaviors such as chewing on a fist, fixating on a feature of the environment and crying when he was not responding. Crying also occurred while he was responding during extinction.

Reconditioning of the response on a FR 5, during session 48, returned the response rate to its previous level during FR 5 reinforcement.

Vocalization

Vocalization was reconditioned in sessions 49 through 51 as shown in the cumulative response records presented in Figure 3. The response rates, 111 responses in CRF, 168 responses in the first FR 3 session and 240 responses in the second, are somewhat higher than those in the prior vocalization sessions under the same schedules. No decline in response rate towards the end of the sessions, as in the prior vocalization sessions, was observed.

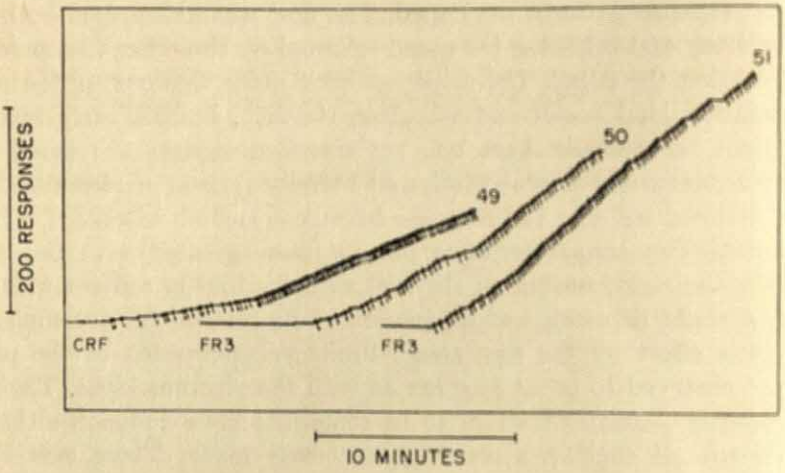


FIG. 3. Cumulative response records of vocalization in sessions 49 (CRF), 50 and 51 (FR 3). Age (60–62 days).

Vocalization and Leg Kick

During sessions 52 and 67 both vocalization and leg kick were studied. Sessions 52 through 58 were devoted to gradually introducing both manipulanda into the experimental situation. In Figure 4 is plotted the number of reinforced (S^D) responses and the number of unreinforced (S^Δ) responses for (A) vocalization plus leg kick, (B) vocalization alone, and (C) leg kick alone. Initially the number of S^D and S^Δ responses were approximately the same for A, B, and C. In sessions 60 and 61 there was no differential responding as the response rate decreased to about one-half its level during session 59 as a result of both responses undergoing extinction during S^Δ. Greater S^D responding is evident starting in session 62 and continuing to increase through session 67. As can be seen in Figure 4 the differential responding developed for both (B) vocalization and (C) leg kick. Differential responding within the five-minute blocks adjusted rapidly when the discriminative stimulus was changed. The percentages of all vocalizations and leg kicks that were S^D responses, all vocalizations that were S^D responses, and all leg kicks that were S^D responses in sessions 59 through 67 can be seen in Table 1. The proportion of S^D responses increases from chance level in the early sessions until about three out of every four responses in the final sessions are S^D responses.

DISCUSSION

This study has presented a demonstration of differential responding under discriminative control for both vocal and motor operants in a

TABLE 1

THE PERCENTAGES OF ALL VOCALIZATIONS AND LEG KICKS THAT WERE S^D RESPONSES, ALL VOCALIZATIONS THAT WERE S^D RESPONSES, AND ALL LEG KICKS THAT WERE S^D RESPONSES IN SESSIONS 59 THROUGH 67

Sessions	Vocalization and leg kick	Vocalization	Leg kick
59	.52	.48	.56
60	.49	.54	.46
61	.49	.59	.41
62	.57	.51	.71
63	.63	.73	.55
64	.67	.65	.68
65	.64	.62	.66
66	.73	.76	.69
67	.72	.69	.77

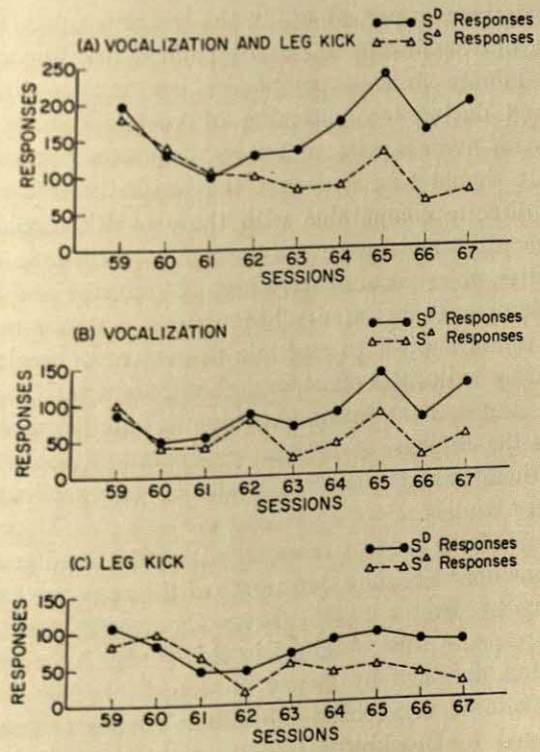


FIG. 4. S^D and S^Δ responses for (A) vocalization and leg kick, (B) vocalization, and (C) leg kick. Sessions 59 through 67. Age (75–83 days).

human infant less than 3 months of age; it has, also, developed an experimental procedure for the analysis of infant operant behavior, identified reinforcers, and achieved schedule control over responding.

This is the first demonstration of differential responding under discriminative control for both vocal and motor operants in a human infant under three months of age. Discriminative control of differential responding in two concurrent responses is the strongest argument that can be advanced to establish that the behaviors under investigation are operant.

Previous research on the conditioning of infant vocal behavior (Rheingold *et al.*, 1959, Weisberg, 1963) have been short-term studies, from 6 to 8 days, with short experimental sessions, from 3 to 10 minutes, and have used 3-month-old infants. By contrast the present study spans 2½ months, from 2 weeks of age to 3 months, and employed experimental sessions of 20- and 30-minutes duration. Both prior studies found an increase in the number of vocalizations during conditioning sessions, using social reinforcements, from about four responses per minute to seven responses per minute in the first study and from one response per minute to three responses per minute in the second study. In the present study the rate of vocalization increased from a baseline of about 1½ responses per minute to seven responses per minute in sessions 12 through 18, and during reconditioning of vocalization in sessions 49 through 51 to an average rate of twelve responses per minute in the final session. It should be noted that the presently obtained response rates are not directly comparable with those of Rheingold *et al.*, and Weisberg since all sounds were automatically recorded, while in the latter two studies, voiced sounds were mainly recorded and reinforced.

The leg-kick operant in infants has a great deal to recommend it. The response is not subject to multiple sources of control; infants can perform it rapidly with little effort, and no apparent fatigue; the operant level is above zero and the maximum response rate is quite high; and a transducer for the response is extremely simple and easily constructed. Given all of these characteristics it would be an excellent response to utilize in future studies.

A comparison of the present research with prior studies of the conditioning of motor operants shows, in terms of the response rates obtained, that the study by Bower (1966) bears the closest resemblance. The average inter-response time obtained in this study, 3 seconds, is slightly higher than that obtained by Bower, 2 seconds, in his study of visual perception in infants using head movement as the response. The procedure employed in Brackbill's (1958) study limited the increase in

response rate to from 2½ responses in 5 minutes to five responses in 5 minutes.

Siqueland and Lipsitt (1966) demonstrated operant control over head movement with infants under 4 days of age. They achieved differential responding under discriminative control using a discrete trial procedure, in which responses to a tactile stimulus were differentially reinforced in the presence of two auditory stimuli. Since a discrete trial procedure was employed no rate measure was obtained.

The development of a successful experimental procedure for the analysis of infant operant behavior was arrived at only after several other procedures were employed and rejected. The problems encountered with the first multiple schedule were avoided through the use of two concurrent multiple schedules. The development of crying during the MULT FR 3 DRO as a result of the occurrence of unreinforced responses, although interesting in itself and worthy of investigation, completely thwarted early attempts to control operant vocalization. Infant crying as a result of unreinforced responding has been reported in other studies (Rheingold *et al.*, 1959; Brackbill, 1958; Heid, 1965) and appears to be a general finding. This suggests that it would be wise to avoid the use of extinction sessions with infants when studying vocal behavior, unless one is specifically interested in the elicitation of crying under an extinction procedure.

The development of crying during the MULT FR 3 NONCONT REIN appeared to be related to the high rate of responding that occurred, although it should not be concluded that there is a definite relationship, since a high rate of vocalization was maintained in sessions 7 through 18, see Figure 1, without the development crying.

For studies whose primary purpose is to provide a demonstration that specific responses of an infant are subject to operant control the use of two concurrent responses, as in this study, is highly recommended. The chief limitation for the usefulness of this procedure is the possibility that the conditions might promote the emergence of superstitious behavior.

We cannot tell from the data how age dependent were the failures and successes of this study. Because of the variation in the procedures employed, we do not know if conditions used late in the study would have been effective earlier; nor do we know if those conditions used earlier would have been effective if the infant had been older. The study of more infants within this age range would clarify this.

A facet of the procedure not to be overlooked or minimized in importance was the location and timing of experimental sessions. A study of this nature would have been extremely difficult to conduct in any

location other than the infant's home. The necessity of having the infant contented (i.e., free from noxious eliciting stimuli) and alert has previously been emphasized. It required careful observation for the entire study and constant attention and care prior to an experimental session to satisfy these requirements. Even then, it was not always possible to satisfy the requirements every day or to predict correctly that they were satisfied.

There were four reinforcers employed in this research. The first, a recording of the mother's heartbeat played through a speaker in the crib, was only used once. Salk (1966) found that a similar recording when played continuously in the nursery had a quieting effect on newborns. It seemed plausible, therefore, that a recording of heartbeat played for 5 seconds might function as a reinforcing stimulus. It was introduced in session 6 on CRF. Following the first response the heartbeat began, the infant appeared unusually alert while it continued for 5 seconds, then when it terminated the infant emitted a scream that was quite unlike any it had produced before; this in turn started the heartbeat again and the infant was quiet until it terminated at which point he screamed. This pattern continued until the session was stopped and the infant removed. The possible reinforcing effects of the heartbeat were overwhelmed, perhaps, by the aversiveness of its termination.

The second reinforcer, the vibrator, appeared to be quite effective, the only exception being the decline in response rate observed at the end of some of the sessions between 7 and 18, which was perhaps attributable to satiation with the reinforcer. The fact that the decline could also be the result of fatigue led to the reduction of the duration of the sessions from 30 minutes to 20-minutes.

The vibrator was replaced by the flashing lights and the recording of the mother's voice in the continuing search for a strong reinforcer. This combination proved to be so effective that it was used for the remainder of the study. The strength of the mother's voice as a secondary reinforcer may have been maintained by having the mother repeat the same monologue on the tape to the infant during every feeding. The effectiveness of the lights alone or the mother's voice alone was not studied.

The success of the procedure was also dependent upon the choice of a suitable discriminative stimulus. To be most effective, a discriminative stimulus in studies with infants should not require an orienting response by the infant and the stimulus ought to affect a different sensory modality than the reinforcing stimulus. A tone, if an auditory reinforcer is not being used, or a pervasive light, as in the present study, are reasonable choices.

REFERENCES

BOWER, T. G. R. Slant perception and shape constance in infants. *Science,* 1966, 151, 832–834.

BRACKBILL, Y. Extinction of the smiling response in infants as a function of reinforcement schedule. *Child Development,* 1958, 29, 115–124.

FERSTER, C. B., AND SKINNER, B. F. *Schedules of reinforcement.* New York. Appleton-Century-Crofts, 1957.

HEID, H. L. An operant analysis of behavior of an infant (from 10 days to 4 months). Unpublished doctoral dissertation, University of Washington, 1965.

RHEINGOLD, H. L., GEWIRTZ, J. L., AND ROSS, H. W. Social conditioning of vocalization in the infant. *Journal of Comparative and Physiological Psychology,* 1959, 52, 68–73.

SALK, L. Thoughts on the concept of imprinting and its place in early human development. *Canadian Psychiatric Association Journal,* 1966, 11, 295–305.

SIQUELAND, E. R., AND LIPSITT, L. P. Conditioned head-turning in human newborns. *Journal of Experimental Child Psychology,* 1966, 3, 356–376.

SKINNER, B. F. *The behavior of organisms: an experimental analysis.* New York: Appleton-Century-Crofts, 1938.

WEISBERG, P. Social and nonsocial conditioning of infant vocalizations. *Child Development,* 1963, 34, 377–388.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 52-69 (1969)

Response of the Human Infant to Level of Complexity of Intermittent Visual Movement[1]

MARSHALL M. HAITH,[2] WILLIAM KESSEN, AND DORIS COLLINS

Yale University

Three levels of visual stimulus complexity were presented to three independent groups of 12 2- to 4-month-old infants. The stimulus was a light which changed position 32 times in a 24-second experimental (E) trial. Three complexity levels were created by varying the predictability of the direction of light movement.

Six 24-second E trials alternated with seven 24-second control (C) trials. *S* was shown the same movement pattern on the first three E trials. On the last three E trials only four *S*s in each of the three complexity groups continued to see the same pattern. The remaining *S*s in each group were shifted to one of the two levels of stimulus complexity not yet experienced.

The dependent variables under observation were limb movement and sucking frequency. In the preshift phase of the experiment, limb movement was suppressed by the most simple and by the most complex levels of stimulation but was facilitated by the intermediate level. The response of hands and feet, although concordant over levels of some variables (e.g., preshift complexity) did not always correspond. There was a suggestion in the later phase of the experiment that a derived score representing a pattern of limb suppression was sensitive to changes in complexity level. Sucking behavior was fairly consistently suppressed by stimulation throughout the experiment. No systematic habituation of responsiveness was found.

Several investigators believe that the increasing ability of the child to organize stimulus information constitutes an important part of perceptual, cognitive, and intellectual development (Bruner, Olver, and Greenfield, 1966; Hebb, 1949; Kagan and Lewis, 1965; Miller, 1956;

[1] The present investigation and the analysis and preparation of this paper were supported by a U. S. Public Health Grant HD-0890 to William Kessen, by a grant from the Carnegie Corporation of New York to William Kessen, and by a U. S. Public Health Grant HD-2680 to Marshall Haith. The actual research was carried out while the senior author was a USPHS Postdoctoral Fellow at Yale University. Mrs. Carol Belove, Mr. Peter Gross, Mrs. Margaret Moore, Miss Karen Stoutsenberger and Mrs. Margaret Yarsawich assisted in the scoring of the data. Special gratitude is owed all the mothers who cooperated in the study. Portions of this report were read at the American Psychological Association meetings in New York, September, 1966.

[2] Now at Harvard University.

Munsinger and Kessen, 1966). A central internal process is postulated which organizes many stimulus events into categories whether the resulting entity be called a "schema," a "structure," or a "chunk." Organization has been thought of as a kind of encoding process which enables the subject to perceive and to respond to whole configurations of stimulus elements or events rather than to each particular element in turn. For instance, the child may encode the stationary or moving configuration of eyes, ears, a nose, lips, and mouth by constructing a "face" category. This process of construction relieves the child of the requirement to respond to every detail of his environment.

It has further been assumed that the child tends to process stimulus elements which are neither too familiar nor too unfamiliar. That is, cognitive change occurs to an incomplete (rather than absent) organization of the stimulus input. If the input has been completely coded it will produce boredom. If the input is too discrepant from any existing organization it will either be feared or ignored. The typical prediction from these ideas has been that there exists an optimal level of stimulus disorganization or variability which will attract and engage the child, thus furthering his development. Two properties of people are commonly considered to be relevant to the construction of categories. The first property is the amount of stimulus information the person has already organized; the second property is the range of discrepancy from his present organization which is optimal for cognitive change.

One problem in testing the foregoing propositions has been the difficulty of defining a dimension of stimulation which also describes a continuum of organization. A dimension of stimulation commonly proposed has been stimulus complexity (Attneave, 1954; Berlyne, 1960; Dember and Earl, 1957). Although there is some suggestion that intermediate amounts of stimulus complexity are preferred to greater or lesser complexity, investigations have not produced uniform results (Hershenson, 1964; Munsinger and Kessen, 1966; Spears, 1964; Thomas, 1965). The most frequent indication of stimulus preference or involvement has been visual attention. (Berlyne, 1958; Cohen, 1966; Fantz, 1958, 1967; Hershenson, Munsinger, and Kessen, 1965).

The present study was carried out to investigate the impact of three levels of stimulus complexity—i.e., predictability of patterns of lights—on limb movement and sucking. Moreover, we wanted to examine how the response to a particular level of complexity would be affected by prior exposure to a different level. That is, we asked what the infant's response would be to a shift, up or down, in the level of complexity presented.

Kagan and Lewis (1965) showed patterns of lights varying in number,

complexity, and area of stimulation to 2- to 4-month-old infants and found no clear effects of these patterns on heartrate and looking time. Cohen (1966) also used patterns of blinking lights as stimuli for infants and found a tendency for them initially to look longer at patterns which showed an intermediate number of position changes.

The manipulation of complexity chosen for the present study was the predictability of directional shifts in a sequence of lights because (1) it was felt that stimulus movement would capture the infant's attention and elicit active ocular involvement, and (2) the use of a movement sequence permitted holding constant the number of stimulus elements presented, their order of presentation, the rate at which the elements were presented, the area of stimulation, and the distance of a shift from one element to the next, while varying only the predictability of the direction of movement.

METHOD

Subjects. The *S*s were 61 human infants 2–4 months of age who lived in the Greater New Haven, Connecticut area. Mothers found out about the study in three ways. A feature story or a short advertisement appeared in the local newspaper, a letter was sent to mothers whose birth announcements appeared in the local newspaper, and instructors, who conducted classes on infant care, told expectant mothers about the study. Mothers were paid only for taxi fare and for babysitting fees for older children. Observations on several *S*s were not completed because of equipment failures ($N = 5$) or because infants rejected the pacifier which was used in the experiment ($N = 9$), cried ($N = 9$), or fell asleep ($N = 2$). The remaining 36 *S*s were assigned to one of nine groups in the present experiment.

Equipment. An air-conditioned observation room in the Institute of Human Relations at Yale housed a modified Skinner baby crib. One-way windows were mounted on two walls of the room to permit observations of the experiment from adjoining sound-insulated alcoves. Recording and switching equipment were kept in one of these alcoves; the mother observed the experiment from the second.

The modified Skinner crib was used to hold a mattress, a head rest, a motion-picture camera, a stimulus indicator box, a stimulus panel, and the baby (Fig. 1). The roof and ends of the crib were 24 × 18-inch interchangeable plywood panels; a plexiglass panel on one side was fixed, a plexiglass panel on the other side could be slipped up and down.

A black masonite stimulus panel, 12 × 12 × 2½ inches, housed 36 Dialco lightbulb sockets in the form of a 6 × 6 matrix. Only nine sockets,

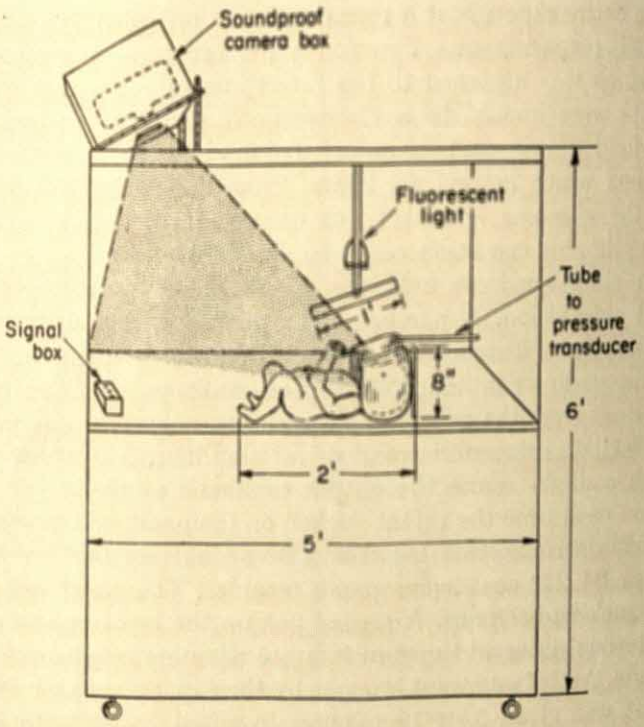


FIG. 1. Observation crib used in the study. The camera field is outlined approximately by dashed lines.

in the form of a 3×3 array, 4×4 inches, were used in the present study. Adjacent lights were 2 inches apart. The nine sockets housed 28-volt bulbs (GE 1819) covered by Dialco 26410-1111 plain red-frosted jewels. A white screen in front of the panel hid the nonilluminated lights. The bulbs on the stimulus panel could be sequentially illuminated by a transistorized ring counter which was silent and which had an interstep dead time of less than 1 millisecond. Associated timing circuitry was used to start and stop the ring counter at appropriate intervals. A combined 14-VDC and 28-VDC regulated power supply was used to supply the switching apparatus, the lights, and an intertrial-interval timer. The 2% regulating efficiency of the power supply insured against dimming of the stimulus lights as a function of line-voltage fluctuations and load variations caused by the operation of other equipment.

For the purpose of recording arm and leg position of the infant at successive time intervals a motor-driven 16-mm Bell and Howell Model 70DL motion-picture camera equipped with an Angenieux 25-mm f1.4 lens was used. The camera was loaded with Kodak Tri-X negative film.

During the entire experiment a repeating timer operated the camera once each second (exposing one film-frame per second) through a Sodeco solenoid which was attached to the "start" button. The camera and its attachments were housed in a Celotex-lined sound-suppressive wooden box. The field of the camera included the whole baby and a stimulus indicator box which housed six lights. From the pattern of illuminated lights in the stimulus indicator box on the developed film-frame, the prevailing stimulus condition could be determined.

A cloth sling headrest held the baby's head toward midline and eliminated side vision. A pacifier was attached to one of the headrest support arms by a firm rubber tube, within which a smaller tube ran from the nipple to a Statham P23AC pressure transducer. The transducer output voltage from the resulting closed-pressure system was routed to a Model 5P1 Glass preamplifier and driver amplifier. A relay connected in series with a diode across the output terminals of the driver amplifier switched on each time the infant sucked on the pacifier. The closed relay contact made a connection between a 6-volt battery and a pen magnet on a Brush B1 202 continuous paper recorder. The result was a digital record of sucking behavior. A second pen on the recorder was activated by the intertrial-interval timer to indicate stimulus termination or onset.

Procedure. All infants were brought by their mothers to the laboratory. The infant was given 5 or 10 minutes to adapt to the unfamiliar surroundings during which time one of the *E*s described the purpose and procedure of the experiment to the mother. Then the mother removed the child's outer clothing leaving the child only in diapers and an undershirt. The mother placed the child on his back on the mattress in the baby crib with his head in the cloth sling headrest. *E* adjusted the pacifier until it seemed comfortably positioned in the mouth of the *S*. All lights were turned off with the exception of a 40-watt fluorescent lamp which directly illuminated the ceiling of the crib and indirectly illuminated the child. The movable plastic side panel was raised a few inches. One *E* stayed in the experimental room out of *S*'s view; the other went to the instrument alcove.

The stimulus panel was 10–12 inches above the child's head, perpendicular to his line of sight. Under these conditions, the stimulus lights occupied approximately 20–24 degrees of the child's visual field. Seven 24-second control (C) trials were presented in alternation with six 24-second experimental (E) trials without interruption. During a C trial only the top center light of the 3×3 bulb array was illuminated. During an E trial the lamps on the stimulus panel were sequentially illuminated, one by one, in accordance with the following rules:

1. Only adjacent bulbs were sequentially illuminated; these were horizontal or vertical to one another (i.e., diagonal movement was not permitted).

2. Only one light was on at a given moment.

3. In each E period 32 position changes occurred at a constant rate of one every 750 milliseconds.

4. All sequences began and ended on the same top-center light. Only one lamp was always lit throughout the experiment. The sequential flashing of lights gave the impression to an adult, not of phi movement, but of a light which intermittently moved and stopped. Within the above constraints three categories of "movement sequences" were generated.

A. *Least complex.* The light made 8 steps (4 cycles/trial) around the perimeter of the matrix in a clockwise or counterclockwise direction.

B. *Intermediate complexity.* The light made 16 steps (2 cycles/trial) in a clockwise or counterclockwise direction but was not constrained to the perimeter of the matrix.

C. *Most complex.* The light made 16 steps (2 cycles/trial) and moved randomly.

Patterns B and C were designed independently for each *S*. Examples of these three levels of complexity of movement sequences are shown in Fig. 2. For a given trial, pattern A was presented four times, while patterns B and C were each presented twice.

The three levels of complexity were each presented to an independent group of twelve *S*s on the first set of three E trials. On the second set of three E trials four *S*s in each group saw the same movement sequence that they had seen on the first three trials. A different four *S*s in each group were shifted to one of the complexity levels that they had not seen earlier. The remaining four *S*s were shifted to the other level of complexity. These subdivisions resulted in nine groups (G) of four *S*s each in the main experiment. Two *S*s in each cell were presented clock-

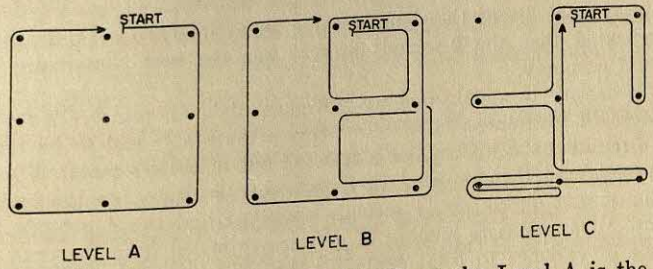


FIG. 2. Examples of light patterns used in the study. Level A is the low-complexity pattern; level B is one of the intermediate-complexity patterns; level C is one of the high-complexity patterns.

wise movement and two *Ss* were presented counterclockwise movement (a variation not relevant to level C) in all E presentations.

RESULTS

Limb Movement

Hand and foot movement were scored by a procedure described by Haith (1966b). Briefly, a single-frame image of the whole baby was projected onto a 6 × 6-foot projection screen. A grid consisting of 120 vertical and 120 horizontal wires stood parallel to the screen, one-half inch in front of it. When one of the wires of the grid was touched by a metal probe, a number representing the coordinate position of the wire was automatically punched into an IBM data card. A scorer touched the probe to each of the four pairs of intersecting wires closest to the hands and feet of the baby thereby recording the coordinate position of each limb. Every other film-frame (corresponding to every other second in real time) was scored.[3] A computer program calculated the change of position (i.e., movement) of each limb between successive frames. The 2-second movement score was the basic datum of analysis.[4]

Each of the seven 24-second C periods and each of the six 24-second E periods was subdivided into four 6-second intervals (I). A mean 2-second movement score was calculated for each I from three 2-second scores. Hand movement and foot movement were computed separately.

Absolute limb movement. A six-way analysis of variance of absolute limb movement on all trials was carried out to evaluate the main factors of groups (G), limbs (L), intervals (I), set (S, preshift vs. postshift), number of presentations (N, Trials 1, 2, and 3 in each set) and treatment (T, E vs. C). All variables, with the exception of G, were within-*S* variables.[5]

The levels of three main factors—Set ($F(1, 27) = 5.53$; $p < .05$),

[3] Preliminary analyses of the data of three *Ss* at intervals of 1, 2, 3, 4, and 5 seconds indicated that the 2-second interval was the best compromise between precision and economy.

[4] It was necessary to supply missing movement data for some cells for some *Ss* either because of instrument failures or because the nipple had to be replaced in the infant's mouth by *E*. When the nipple fell out of the *S*'s mouth *E* waited for about 5 seconds and then replaced it. Data were substituted on the basis of row and column means during the 4 seconds preceding and the 4 seconds following replacement. An appropriate number of *df* was subtracted from the error terms in the analyses of variance to compensate for the substituted scores. If a *S* had the nipple replaced more than three times, his data were not used. The same procedure was followed in substituting missing sucking data.

[5] All within-*S* factors and their interactions in this and subsequent analyses were tested against the appropriate nonpooled *S* × Within-Factor interaction.

Interval ($F(3, 81) = 4.36$; $p < .01$), and Number of presentations ($F(2, 54) = 6.83$; $p < .01$)—were reliably different and the interaction of these three factors ($S \times I \times N$) was also reliable ($F(6, 162) = 2.53$; $p < .05$).

The general trend for all three variables, S, I, and N was an increase in absolute movement over time. The $S \times I \times N$ interaction is plotted in Fig. 3.

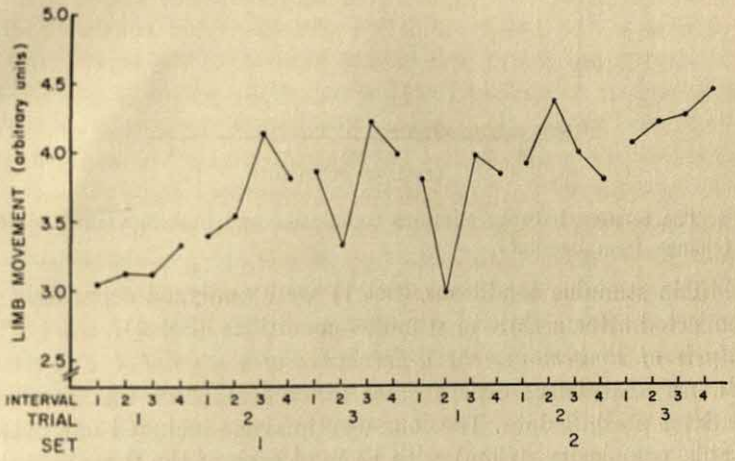


Fig. 3. Absolute limb-movement scores across the course of the study, for all groups, E and C trials combined.

If the secular trend over time is ignored, the finding that emerges is a curvilinear relation between the number of trials and the difference between interval means within a trial. The interval curve is relatively monotonic on Trial 1, becomes somewhat nonmonotonic on Trial 2, reaches a maximum of nonmonotonicity on Trials 3 and 4, and then returns gradually to monotonicity on Trials 5 and 6. Thus, *Ss'* movement increased over the course of the experiment and presumably their sensitivity to stimulus change first increased and then decreased whether to onset or offset.

Movement difference scores. The analyses of absolute movement did not use the information contained in responding during the baseline control immediately before an E period. Therefore, scores for experimental trials were calculated for hand movement and foot movement in the following manner. Each of the four interval scores in an E trial was subtracted from a baseline score. The baseline was the average of the last two interval scores in the C trial which immediately preceded the E trial under consideration. Limb-movement difference scores from two parts of the study were analyzed separately. Data collected prior

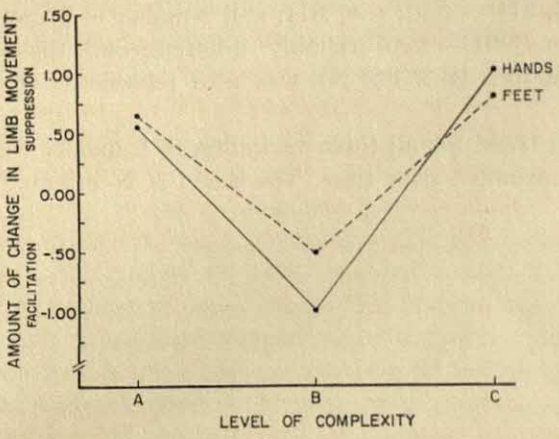


FIG. 4. The relation between stimulus complexity and limb-movement difference scores (change from baseline).

to a shift in stimulus conditions (Set 1) were analyzed separately from data collected after a shift in stimulus conditions (Set 2).

Analysis of limb-movement difference scores on Set 1 experimental periods. An analysis of experimental difference scores was carried out on the Set 1 preshift data. The four-way analysis included one between-*S* variable, complexity, (Com) with 12 *S*s at each of the three complexity levels. The three within-*S* variables were L, N, and I.

The complexity factor ($F(2, 33) = 4.37$; $p < .05$) and the L × N interaction ($F(2, 66) = 5.40$; $p < .01$) were reliable. The difference scores as a function of level of complexity for hand and for foot movement are plotted in Fig. 4. It should be noted that positive scores indicate *suppression* of movement while negative scores indicate *facilitation* or increased movement in E intervals relative to the baseline.

Movement decreased when stimulus sequences A and C were presented. However, movement increased when sequence B was presented (A vs. B, $p < .05$; B vs. C, $p < .01$; A vs. C, n. s.). It is important to note that the baseline levels on which the difference scores for the three complexity groups were based were approximately equal.

Although the limbs did not differ in response to differences in complexity level the plot of the L × N interaction in Fig. 5 indicates that the limbs behaved differently across the three stimulus presentations. A decrease in hand movement did not occur on Trials 1 and 3 but it did occur on Trial 2. Foot movement also was suppressed only a slight amount on Trial 1, but appeared to increase on Trial 2, and then, on Trial 3, was dramatically suppressed.

Analysis of limb-movement difference scores on Set 1 control periods.

An analysis on Set 1 C trials identical to that on Set 1 E trials was carried out to study the effects of stimulus termination. The baseline for C trial difference scores was an average of responding in the preceding 12 seconds of E stimulation. The Com and I factors were reliable. The Com function for C trials ($F(2, 33) = 4.43$; $p < .05$) mirrors that for E trials, perhaps because the baseline score for a C trial was based on the last two intervals in the preceding E trial.

The I function ($F(3, 99) = 5.44$; $p < .01$) for C trials, which was not predictable from the E baseline level, is shown in Fig. 6. One might have expected the interval difference scores for control periods to show a linear trend toward greater movement (facilitation) thereby reflecting the overall trend over time shown in Fig. 3. It can be seen that there is virtually no change during Interval 1, a decrease during Interval 2, and then a return to increased movement during Intervals 3 and 4 (2 vs. 3 or 4, $p < .01$). The stable suppression of movement in Interval 2 suggests that termination of stimulus movement exerts a systematic influence on behavior but requires several seconds before the effect is manifested. For comparison's sake, the I function for E trials, which did not show statistically reliable variation is also shown in Fig. 6.

Although additional analyses on control-trial responding were carried out none revealed systematic effects. Therefore these analyses will not be discussed in further detail.

Analysis of limb-movement difference scores on Set 2 experimental periods. An analysis of variance of limb movement was carried out on E difference scores on the postshift set of trials. The Ss were blocked into three groups on the basis of preshift complexity. The resulting analysis contained five factors—Blocks (Bk), Com, L, T, and N in a

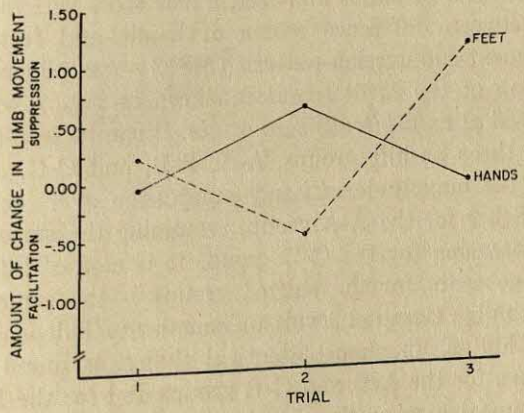


FIG. 5. The relation between Set 1 trial number and limb-movement difference scores (change from baseline).

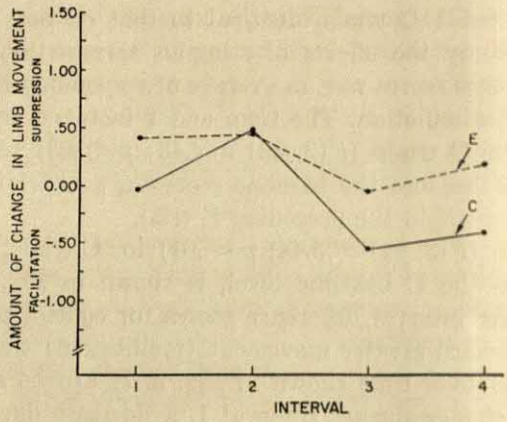


FIG. 6. The relation between intervals, E and C trials of Set 1, and limb-movement difference scores (change from baseline).

$3 \times 3 \times 2 \times 4 \times 3$ mixed design. None of the main factors produced a reliable effect on movement.

A reliable three-way interaction of Bk, Com, and L ($F4, 27 = 4.50$; $p < .01$) suggests that the preshift complexity level affected the response to the postshift complexity level in a functionally different way for hands and feet. Stated in another way, the discrepancy between hand and foot difference scores on the second set of E trials was determined by particular combinations of preshift and postshift complexity (A-A vs. A-B, $p < .01$; A-A vs. B-A, $p < .01$; A-A vs. C-A, $p = .05$; C-C vs. C-A, $p < .01$; C-C vs. C-B, $p < .05$; C-C vs. A-C, $p < .01$; C-C vs. B-C, $p < .05$; A-A vs. C-C, $p < .01$; B-B vs. C-C, $p < .01$; all other comparisons, $p > .05$). To help in a tentative explanation of the discrepancy between the behavior of hands and feet, a new score was calculated. The discrepancy between difference scores of hands and feet (Dhands − Dfeet)—movement-suppression pattern (MSP)—was plotted for all nine groups. A graph of the MSP scores is shown in Fig. 7 with the Set 1 MSP scores *fixed at an arbitrary zero origin*. It can be seen that the MSP scores for the three noshift groups, A-A, B-B, and C-C, are strikingly different with the magnitude of hand suppression over foot suppression increasing in Set 2 for the A-A group, remaining the same for the B-B group, and decreasing for the C-C group. It is also of interest to note that the change score for the shifted groups is the same in size and direction for similar terminal levels of complexity independently of the initial level. That is, an almost identical change is found for the A-C and B-C groups, for the A-B and C-B groups and for the B-A and C-A groups. To summarize, when *Ss* were shifted from one level of complexity to another, only the terminal level determined the size and direction of

the change in the MSP score. When *Ss* were not shifted, the MSP score was a function of complexity level in both Set 1 and Set 2. For reasons that are by no means obvious, A-A *Ss* showed relatively higher hand-movement suppression in Set 2 while C-C *Ss* showed relatively higher foot-movement suppression.

Sucking

As in the case of limb movement, analyses were carried out on sucking during four 6-second intervals for seven C periods and six E periods. The absolute sucking scores represent the average number of sucks during 2 seconds of real time.

Absolute sucking. A 5-way analysis of variance of absolute sucking was carried out to evaluate the main factors of groups (G), intervals (I), set (S), number of presentations (N), and treatment (T). All of these variables, with the exception of G, were within-*S* variables.

The only major factor that was stable was T. E-Period sucking was reliably lower than C-period sucking ($F(1, 27) = 12.57$; $p < .01$). The G × N ($F(16, 54) = 2.20$; $p < .05$) and the S × N × T ($F(2, 54) = 3.72$; $p < .05$) interactions were also reliable. As can be seen in Fig. 8 there was a tendency for sucking to increase over time during control trials. However, for the experimental trials, with the exception of Trial 1 sucking, which is depressed, sucking rate over time remained relatively constant.

Figure 9 shows the change in overall sucking rate with time for E and C trials combined. This graph can be compared with Fig. 3 which shows the same functions for limb movement. It can be seen that the general trend for sucking, as for movement, is toward an increase over time but that the functions are not parallel in detail.

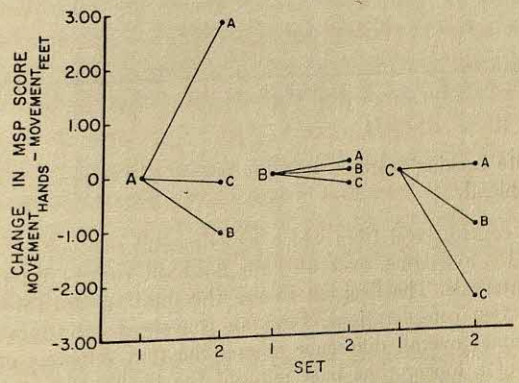


FIG. 7. Movement-suppression pattern (MSP) after shift of stimulation with pre-shift values set arbitrarily at zero.

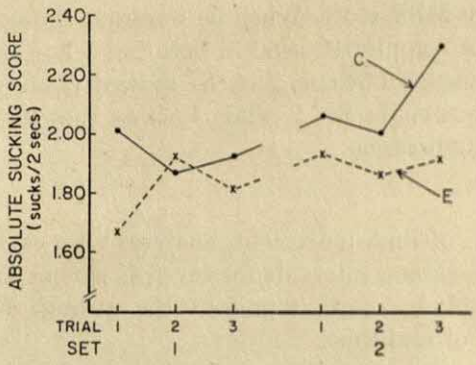


FIG. 8. Absolute sucking scores across the course of the study, for all groups, intervals combined.

Sucking difference scores. The method used for deriving movement difference scores was used to derive sucking difference scores.[6]

Analysis of sucking difference scores on Set 1 experimental trials. An analysis of sucking difference scores for the first set of trials was performed in a $3 \times 4 \times 3$ design with Com, I, and N as the main factors. The only reliable finding was that for N ($F(2, 66 = 3.59; p < .05)$. The greatest suppression of sucking occurred on Trial 1. On Trial 2, there was a slight increment in sucking. Although the general shape of the function relating sucking activity to complexity level was similar to that

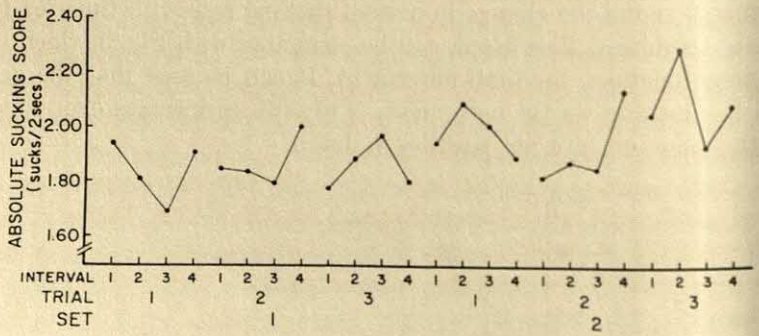


FIG. 9. Absolute sucking scores across the course of the study, for all groups, E and C trials combined.

[6] The baseline control level from which experimental suck scores were subtracted was determined by averaging over all four 6-second control intervals instead of over only two intervals. The decision to use the total control interval as baseline was dictated by two considerations. First, in contrast to the movement measure, there was a minimal overall difference among the four intervals on control trials. Second, the use of a longer time base reduced the likelihood of spurious baselines which might have been produced by sampling during a "burst" or a "rest-between-bursts," rhythmic characteristics of sucking behavior.

relating limb activity to complexity, the differences for sucking were not statistically stable.

Analysis of sucking difference scores on Set 2 experimental periods. A four-way analysis of variance on the main factors of Bk, Com, I, and N, constituting a $3 \times 3 \times 4 \times 3$ design was carried out on the second set of E trials. One major factor, Com, ($F(2, 27) = 3.57$; $p < .05$) and the Bk $\times$ Com $\times$ N ($F(8, 54) = 2.88$; $p < .05$) interaction were significant. The means for the A, B, and C levels were .00, .20, and .40, respectively (A vs. C, $p < .05$).

Correlations Among Response Measures

Frequency histograms of absolute activity scores in a 6-second interval revealed a negatively skewed distribution both for hand and for foot movement. The distribution for sucking was almost rectangular. In contrast to the histograms of absolute scores, however, frequency histograms for *difference* scores showed a virtually normal distribution for hands, feet, and sucking.

The stability of the dependent variables was reflected in the correlation (r) between Set 1 and Set 2 of *baseline* responding for hands, feet and sucking; .68, .86, and .81, respectively.

The usefulness of difference scores is related directly to the correlation between absolute baseline and absolute E-period scores. The correlations between baseline and E-period interval scores in Set 1 for hand movement ranged from .17 to .61 with a median correlation of .38. The correlations for foot movement ranged from .46 to .73 with a median correlation of .55. The correlations for sucking ranged from .36 to .71 with a median correlation of .55. The .05 and .01 levels of significance equal correlations of .27 and .38, respectively.

The relationship among response measures was determined by correlating combined baseline scores *between* Ss, for hands, feet, and sucks. The correlations for hand movement vs. foot movement, hand movement vs. sucking, and foot movement vs. sucking were .72, −.63, and −.63, respectively.[7]

The correlations among response measures assessed *between* Ss do not reflect whether or not a *change* in one variable was accompanied by a *change* in the other for a particular S (Mandler, 1959). In contrast to the between-S correlations, intraclass correlations between the hand, foot, and sucking variables showed no general moment-to-moment relation for absolute, difference or MSP activity scores.

[7] It should be noted once again that sucking baseline scores were calculated on 24 seconds of C time while limb movement scores were calculated on 12 seconds of C time. However, the sucking intercorrelations were actually slightly higher for the longer time base.

A check was made on tendencies for the right or left limbs to be more active across all trials and on tendencies for the more active arm and leg to be homolateral. The number of *Ss* in each of the four possible patterns of laterality was virtually equal.

DISCUSSION

The principal findings of the present study may be summarized in the following way. During preshift stimulation, limb activity was suppressed by the most simple and most complex patterns of visual movement but was facilitated by the pattern of intermediate complexity. The same nonlinear pattern was found between sucking and complexity but sucking was consistently suppressed by stimulation and the obtained differences were not statistically stable. The behavior of hands and feet, although concordant over levels of some variables (e.g., preshift complexity), did not always correspond. In the postshift phase of the experiment there was a suggestion that a score representing the difference in suppression between hands and feet was sensitive to changes in complexity level. No systematic habituation of responsiveness, measured by limb or sucking activity, was found. The one exception to this finding was a large suppression of sucking to stimulation on the first experimental trial followed by smaller but virtually equal suppressions on the remaining trials.

The present findings are at variance with those of Kagan and Lewis (1965). Kagan and Lewis reported arm movement in response to three levels of complexity of flashing lights. Each of their patterns was presented with a 1-second interval between successive flashes. No differential effect of complexity level was found but "habituation" as measured by increased arm movement as a function of repeated stimulus presentations was indicated. Among the many differences between the present study and that by Kagan and Lewis the present study used no interflash interval. It is possible that the response to alternating offset and onset of light in their study masked any differential response to complexity level. It is important to note that their investigation did not utilize control intervals to provide for the possibility that general restlessness of their *Ss* increased over time rather than "habituation" occurring. The present study clearly favors the former interpretation inasmuch as absolute movement increased significantly over time but the difference between absolute movement during C trials and that during E trials did not decrease.

In the present study the predictability of direction of light movement was the only stimulus parameter varied over the three levels of complexity while the area of stimulation, the distance between successive

light positions, and the number of position changes were held constant. Within the above constraints the full range of complexity was covered.

If suppression of ongoing activity is regarded as an index of attention (Berlyne, 1960), then an optimal-level hypothesis would lead to the prediction of an inverted-U relation between suppression of activity and stimulus complexity. Exactly the opposite curvilinear function was found here—suppression of activity to Levels A (low complexity) and C (high complexity) and increase of activity to intermediate Level B. However, it is important to remember that the present study manipulated complexity by varying *sequential* characteristics of visual stimulation as opposed to more conventional procedures which have manipulated simultaneous characteristics of visual stimulation, e.g., characteristics of geometric forms.

The seemingly paradoxical effect of complexity or relative predictability of stimulus movement) on activity may be better understood if one considers the possibility that Ss, during an E trial, could have repeatedly formed an expectancy for the next direction of movement of the light. For instance, after one or two cycles of the A pattern the infants may have begun to anticipate the next positional shift of the light. An expectancy would presumably establish a readiness in *S* to make a particular eye movement which would be confirmed or disconfirmed by the actual light movement. The ratio of disconfirmations to confirmations as well as frustration resulting from wrong anticipations are assumed to increase as complexity increased.

Assume that a decrement in activity reflected attention and an increment in activity reflected frustration. Then lowered activity during level A stimulation would accompany attention and active visual tracking of the stimulus with little frustration from disconfirmation. Level B was also attended and tracked but frustration, resulting from an increased number of disconfirmations, manifested itself in increased activity. Strong attention to Level C stimulation is suggested by the occurrence of decreased activity. However, by the arguments advanced earlier an increase in activity would be expected from a higher number of disconfirmations and the resulting higher level of frustration. This theoretical objection is resolved if either Ss did *not* visually track the stimulus or they tracked only on an *ad hoc* basis; that is, the stimulus could have been tracked, point for point, but the tendency to establish expectancies might either have been extinguished very early or never have been established because Level C was so unpredictable. Alternative interpretations are possible. But any interpretation which explains differential responding to sequential stimuli differing only in predictability of position, will almost certainly depend either on notions of expectancy or

on *S*'s matching of current and past visual and oculomotor events. The interpretation presented here is favored because it is testable.

The finding that hand and foot difference scores were often of different algebraic signs, although somewhat puzzling, is interesting in the light of traditional stabilimetric methods of measuring activity in infants (Irwin and Weiss, 1934; Rovee and Levin, 1966). Stabilimetric procedures do not permit separate analyses of limb movement. Future studies may profitably consider the *pattern* of the child's limb movements in addition to the absolute level of activity. The Movement-Suppression-Pattern (MSP) scores were used here to attempt to represent the pattern of limb activity. The derived MSP score showed stable differences between shift and no-shift groups both for the same terminal and for the same departure levels of complexity.

Sucking, measured as number of expression movements (Sameroff, 1965), was very consistently suppressed by stimulation. Moreover, the distribution of sucking difference scores was strikingly normal and correlations involving sucking were high. In some respects sucking was an ideal dependent variable. However, for the most part sucking behavior, as measured here, was not discriminative of the functions under study. The one exception to this statement was the positive linear function relating sucking difference scores to the level of second-set complexity. However, even this function was complicated by a perplexing $G \times C \times N$ interaction.

Finally, in the present study, as in an earlier study with newborn infants using intermittent visual movement (Haith, 1966a), suppression of sucking to stimulation was found but no systematic intertrial habituation of responsiveness occurred. It may be that intermittent visual movement was too interesting a stimulus for habituation to occur during the period of observation.

REFERENCES

ATTNEAVE, F. Some informational aspects of visual perception. *Psychological Review*, 1954, **61**, 183–193.

BERLYNE, D. E. The influence of the albedo and complexity of stimuli on visual fixation in the human infant. *British Journal of Psychology*, 1958, **49**, 315–318.

BERLYNE, D. E. *Conflict arousal and curiosity*. New York: McGraw-Hill, 1960.

BRUNER, J. S., OLVER, R. S., AND GREENFIELD, P. M. *Studies in cognitive growth*. New York: Wiley, 1966.

COHEN, L. B. Observing responses, visual preferences, and habituation to visual stimuli in infants. Unpublished doctoral dissertation, University of California, Los Angeles, 1966.

DEMBER, W. N., AND EARL, R. W. Analysis of exploratory, manipulatory, and curiosity behaviors. *Psychological Review*, 1957, **64**, 91–96.

FANTZ, R. L. Pattern vision in young infants. *Psychological Record*, 1958, **8**, 43–47.

FANTZ, R. L. Visual perception and experience in early infancy: A look at the hid-

den side of behavior development. In H. W. Stevenson, E. H. Hess, and H. L. Rheingold (Eds.), *Early behavior: Comparative and developmental approaches.* New York: Wiley, 1967, pp. 181–224.

HAITH, M. M. The response of the human newborn to visual movement. *Journal of Experimental Child Psychology,* 1966, 3, 235–243. (a)

HAITH, M. M. A semiautomatic procedure for measuring changes in position. *Journal of Experimental Child Psychology,* 1966, 3, 289–295. (b)

HEBB, D. O. *Organization of Behavior.* New York: Wiley, 1949.

HERSHENSON, M. Visual discrimination in the human newborn. *Journal of Comparative and Physiological Psychology,* 1964, 58, 270–276.

HERSHENSON, M., MUNSINGER, W., AND KESSEN, W. Preference for shapes of intermediate variability in the human newborn. *Science,* 1965, 147, 630–631.

IRWIN, O. C., AND WEISS, L. A. Differential variations in the activity and crying of the newborn infant under different intensities of light: A comparison of observational with polygraphic findings. *University of Iowa Studies in Child Welfare,* 1934, 9, 137–147.

KAGAN, J., AND LEWIS, M. Studies of attention in the human infant. *Merrill-Palmer Quarterly,* 1965, 11, 95–127.

MANDLER, G. Stimulus variables and subject variables: A caution. *Psychological Review,* 1959, 66, (3), 145–149.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 1956, 63, 81–97.

MUNSINGER, H., AND KESSEN, W. Stimulus variability and cognitive change. *Psychological Review,* 1966, 73, 164–178.

ROVEE, C. K., AND LEVIN, G. R. Oral "pacification" and arousal in the human newborn. *Journal of Experimental Child Psychology,* 1966, 3, 1–17.

SAMEROFF, A. J. An apparatus for recording sucking and controlling feeding in the first days of life. *Psychonomic Science,* 1965, 2, 355–356.

SPEARS, W. C. Assessment of visual preference and discrimination in the four-month-old infant. *Journal of Comparative and Physiological Psychology,* 1964, 57, 381–386.

THOMAS, H. Visual-fixation responses of infants to stimuli of varying complexity. *Child Development,* 1965, 36, 629–638.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 70–80 (1969)

The Acquisition and Violation of Expectancy: An Experimental Paradigm[1]

MICHAEL LEWIS
Educational Testing Service

AND

SUSAN GOLDBERG
University of Zambia

An experiment, using 44-month-old children, was conducted wherein a repeated presentation of a stimulus (S_1) was followed by a new stimulus (S_2). It was predicted that attention, as measured by orientation of the head and fixation of the eyes, cardiac deceleration, smiling, and pointing, would show response decrement over repeated trials of S_1 and would show response recovery to S_2. The data of this experiment confirmed these hypotheses and the results are discussed in terms of the orienting reflex, emotion, and general theories of attentional processes in the young child.

The dimension of novelty and familiarity is an important stimulus property influencing attentional behavior (Berlyne, 1960; Sokolov, 1963; Cantor, 1963). One definition of novelty and familiarity is based on the assumed frequency with which certain stimuli have occurred in the organism's experiential history (Lewis, 1965). This definition usually limits the experimenter to choosing gross distortions of reality as examples of novel stimuli. Novelty and familiarity can also be defined experimentally by controlling the frequencies with which stimuli occur. A stimulus S_1 is defined as familiar when it has been presented repeatedly for n trials. The occurrence of a second stimulus, S_2, on trial $n + 1$ defines S_2 as a novel event. S_2 may be any event discriminable from S_1. Both definitions involve the notion of confirmation and violation of expectations.

Sokolov (1963) has used this paradigm to investigate the orienting reflex (OR). Indeed, he has defined the OR as those responses which

[1] This research was conducted at the Fels Research Institute and was supported in part by grants HD-00868, FR-00222, and FR-05537 from the National Institute of Mental Health, United States Public Health Service. The authors wish to thank Helen Campbell who served as one of the experimenters.

habituate to a repeated S_1 and which recover when S_1 is altered. Sokolov argues that this response decrement and recovery are mediated by some central process such as memory or neuronal model formation. When there is no match between the external event and the internal model, central excitation occurs resulting in orienting behavior. When there is a match, central inhibition occurs and little orienting behavior results. Within this theoretical system, rate of response decrement can be viewed as a function of the speed of model acquisition whereas response recovery is a function of discriminability, model formation, and the nature and degree of the violation (S_2—distortion, change in complexity, form, shape, color). Thus, novelty and familiarity can be defined operationally by the same experimental manipulations as the orienting reflex.

In a preliminary study, Lewis, Goldberg, and Rausch (1967) concluded that visual attention in the preschool child, as measured by total fixation (TF), can be considered part of the orienting reflex. The authors found response decrement to repeated presentation of S_1 and recovery to a variant of S_1. In addition to the TF data of the Lewis *et al.* study, behavioral observations indicated that *S*s showed increased smiling, pointing, and expressions of surprise when the stimulus was changed. This suggests that altering S_1 may result in affective responses. Moreover, since autonomic feedback to central neural structures affects stimulus reception, and therefore, model formation (Sokolov, 1963), such autonomic measures as cardiac response should be associated with the orienting reflex.

The experiment reported here was designed both to replicate the earlier finding and to further investigate the effect of experimental manipulation of novelty and familiarity on visual attention and its physiological and behavioral correlates.

METHOD

Subjects

*S*s were 20 children (10 boys and 10 girls) 3½ years of age, who saw the film presentation as part of a longitudinal study of attention in which they were subjects. Each *S* was brought to the laboratory in the morning by his mother.

Equipment and Procedure

Each *S* was seated at a table enclosed in a uniform gray room approximately 5 × 5 feet. The child's mother sat to the rear and side of *S*. Visual stimuli were presented by rear-screen projection approximately 2½ feet from *S*'s head and at *S*'s eye level.

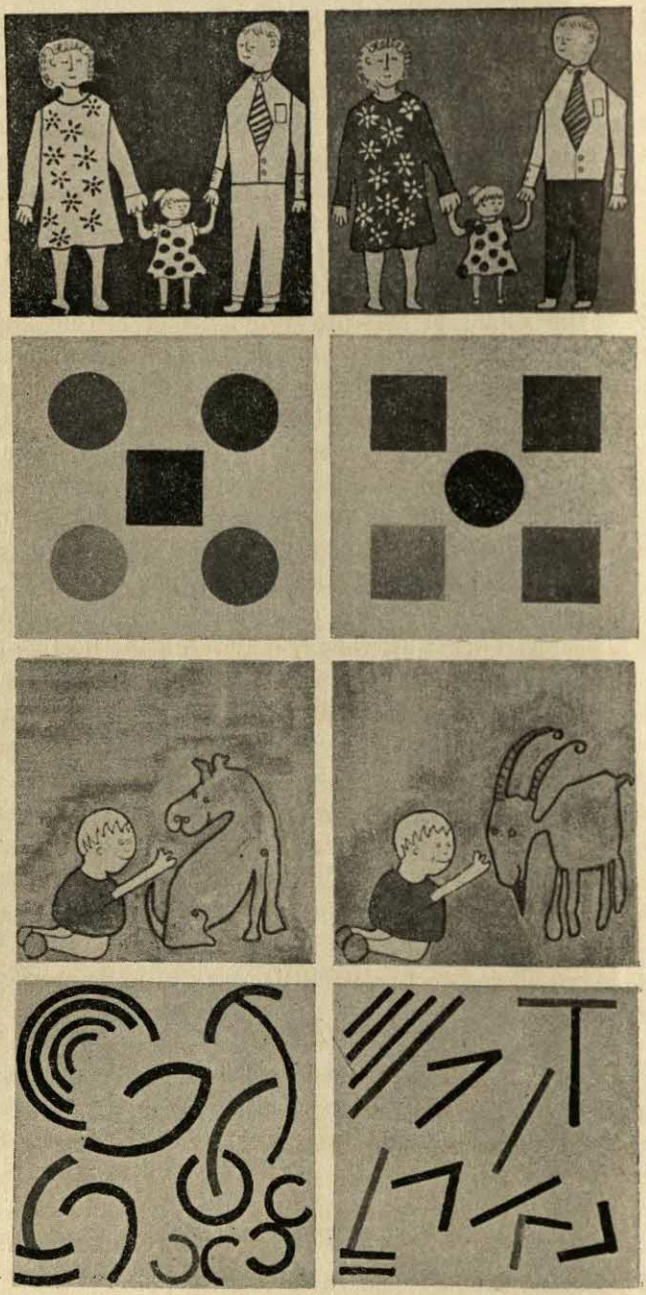


FIG. 1. Visual stimuli shown in sets from A (top) to D (bottom). S_1 is shown on the left, S_2 on the right for each set. All stimuli except S_1 in set A are chromatic. For set A, the violation was the change from an achromatic to a chromatic picture; for set B, the violation was a change in form; for set C, content; and for set D, curvature.

Four different sets of stimuli were presented (Fig. 1) and the order of presentation was the same for each *S*. Set A was presented first, followed by sets B, C, and D. In each set, the stimulus on the left was presented six times for 30 seconds with a 30-second intertrial interval. On trial 7, the stimulus on the right was presented for 30 seconds. After an intertrial interval of 30 seconds, the next set began immediately. All 20 *S*s received this presentation sequence.

Measures of Attention and Measurement Procedures

Measures of attention. During the film presentation, visual orientation to the screen was recorded by two observers (interscorer reliability = .94) who were unaware of which stimulus was on the screen. Each observer depressed a key each time *S*'s eyes were oriented toward the screen and released it when *S* turned away. The state of the observer's key was automatically recorded on an event recorder along with stimulus onset and offset. Smiling, pointing, and surprise were recorded in similar fashion by two other observers. The occurrence of surprise was based upon the observers' judgment of *S*'s facial expression, and spontaneous verbalizations, e.g., "Oh gee! Look at the new picture." Reliability of behavioral measures recorded under these conditions (each observer recording smiling, pointing, and surprise simultaneously) ranges from .68 to .96. Cardiac response was recorded both on a punched paper tape (see Welford, 1962) and a polygraph. The Fels cardiotachometer[2] converted each r-r interval of the ECG to a rate in beats per minute and a continuous beat-to-beat cardiac rate was recorded on the polygraph. The measure of cardiac response was the difference between the mean of the three beats preceding stimulus onset and the three lowest beats occurring during fixation in the stimulus period. If a subject did not look at the stimulus or did not look long enough to record three heart beats, there was no cardiac response recorded for that trial.

Measurement procedures. In the study of response decrement and recovery, a variety of measurement procedures are possible. For response decrement, one easily calculated measure is the difference between response on trial 1 and response on trial 6. The trial 1 minus 6 measure of response decrement has the advantage of being straightforward and easy to calculate and manipulate. However, it does not utilize data from intermediate trials. A method that does utilize the data is a regression analysis. The slope of a linear regression and the exponent of an exponential represent rate of response decrement. Response recovery also can be measured by several different procedures. For ex-

[2] Fels Cardiotachometer is manufactured by the Yellow Springs Instrument Company, Yellow Springs, Ohio 45387.

ample, fixation on trial 7 can be compared to trial 6, or an experimental and control group can be used when the control group receives S_1 on trial 7 and the experimental group receives S_2. The difference between control and experimental groups on trial 7 would reflect response recovery. In the preliminary study by Lewis *et al.* (1967), this latter procedure was followed. The results showed a significant difference between the experimental and control groups and indicated that altering S_1 produced response recovery. An alternative method of evaluating response recovery is to compare the observed response on trial 7 with that predicted by the regression equation. This has the advantage of utilizing all data for trials 1–6 in predicting and evaluating observed behavior on trial 7. In effect, each group serves as its own control with behavior on trial 7 evaluated on the basis of parameters of previous behavior by the same *S*s. This is the analysis which will be used in this paper. Although this technique eliminates the need for a separate control group, data from the control group of the preliminary study will be presented for comparison, when available, in order to demonstrate that the results generated by the two methods are similar.

RESULTS

Both linear and curvilinear regression analyses were carried out for each variable for sets A, B, C, D, and the mean over all sets. The curvilinear function was of the form $Y = A + Be^{-Cx}$ for values of C from .25 to 2.5. The negative exponential generally fit the data better than the linear equation. Since Lewis *et al.* reported no significant differences between sets and the present data confirmed this, the following discussion will focus primarily upon the mean curve over all sets.

Total fixation. Figure 2A shows the mean TF over all four sets and indicates the best fit regression curve, $Y = 10.67 + 19.98e^{-.5x}$ ($F = 60.75$, $p < .01$). In all sets, the negative exponential accounted for a significant proportion of the variance and gave a better fit than the best linear function. The raw data points indicate rapid response decrement over trials 1–6.

The observed TF on trial 7, as shown in Fig. 2A, represents a distance of 7.21 standard deviations from the point predicted by the regression curve ($p < 1 \times 10^{-5}$) indicating response recovery to S_2. In the preliminary study, the control groups observed TF on trial 7 was only 0.92 standard deviations from the predicted value, a nonsignificant difference.

Cardiac response. Cardiac response was measured in terms of number of beats deceleration from base to stimulus period. Figure 2B presents the mean number of beats deceleration over all four sets for each of the seven trials. The initial response to each set was an average deceleration

of 10–14 beats per minute. With repeated presentation over trial 1–6, there was a decrease in the amount of deceleration. These data give a significant negative exponential fit, $Y = 4.58 + 11.07e^{-.5x}$ ($F = 20.23$, $p < .05$).

The response to the violation on trial 7 was increased deceleration. The observed deceleration on trial 7 (Fig. 2B) is 3.62 standard deviations above the prediction based on the regression curve ($p < 2 \times 10^{-4}$). Thus, cardiac deceleration parallels the fixation response, decreasing to repeated S_1 and increasing when S_1 is altered. The correlation between fixation (TF) and cardiac deceleration was $r = .70$ ($p < .01$) over the 28 trials of the experimental session.

Smiling response. The mean number of seconds smiling over all four sets is shown in Fig. 3A. There was a consistent decrease in amount of smiling over trials 1–6 and the best fit curve for the data points for each set was a negative exponential. The regression equation for the mean over all sets was $Y = 1.40 + 8.80e^{-x}$ ($F = 72.27$, $p < .01$).

The observed amount of smiling on trial 7 as shown in Fig. 3A was 6.64 standard deviations above the regression curve ($p < 3 \times 10^{-5}$). Smiling data were also available for the control group in the earlier

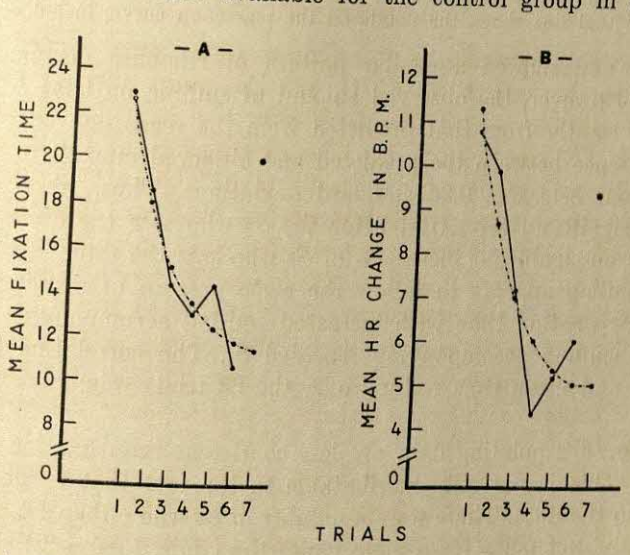


FIG. 2. A. Mean total fixation time in seconds for each trial averaged over sets A–D. B. Mean heart rate change in beats per minute for each trial averaged over sets A–D. Since the cardiac response was a deceleration in heart rate, the Y-axis indicates number of beats decrease with larger numbers indicating greater decelerations. In both graphs, the solid line indicates observed points with trial 7 (violation) shown as an isolated dot. The dotted line represents the points predicted by the regression equation including trial 7.

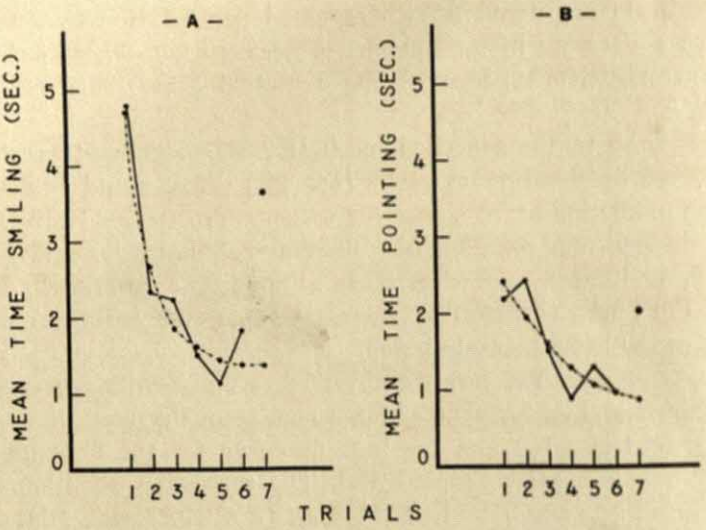


FIG. 3. A. Mean time smiling in seconds for each trial averaged over sets A–D. B. Mean time pointing in seconds for each trial averaged over sets A–D. In both graphs, the solid line indicates observed data with trial 7 shown as an isolated point. The dotted line shows the points on the regression curve, including trial 7.

experiment and showed a similar pattern of response decrement for trials 1–6. However, the observed amount of smiling on trial 7 did not differ significantly from that predicted from the regression curves—the mean difference between the predicted and observed amount of smiling over all four sets was 0.26 standard deviations. Thus, while smiling increased significantly on trial 7 for the *Ss* who saw the altered trial, there were no significant increases for *Ss* who saw the same stimulus on trial 7. Smiling appears to follow the same pattern of decrement and recovery as fixation time with increased smiling accompanying S_2 and decreased smiling accompanying repeated S_1. The correlation between mean TF and mean time smiling over the 28 trials was: $r = .66$ ($p < .01$).

Pointing. The pointing data are less consistent than fixation or smiling data. This is partially attributable to the fact that fewer *Ss* contributed to the data. There were a number of *Ss* who either did not point at all or did not point for a majority of the trials. Figure 3B shows the mean time pointing over all four sets for each trial and the data again fit a negative exponential, $Y = .27 + 2.74e^{-.25x}$ ($F = 11.80$, $p < .05$).

The observed amount of pointing for trial 7 was significantly above the regression line for each of the four sets. The mean amount of pointing over all four sets on trial 7 was 3.65 standard deviations above the predicted point ($p < 2 \times 10^{-4}$). Similar data for the control *Ss* in the

preliminary study indicated a decrease in the amount of pointing for trials 1–6 and the mean pointing time on trial 7 showed continued decrease in response—the mean difference between the predicted and observed amount of pointing over all four sets was only 0.26 standard deviations. Thus, like smiling, pointing increased on trial 7 only when *Ss* were exposed to the altered stimulus.

Surprise response. There were not enough instances of surprise recorded to permit any statistical analysis. However, 13 of the 14 recorded occurrences of surprise were on trials 1 or 7, that is, the trials on which a new stimulus was introduced.

DISCUSSION

Measures of Response Decrement and Recovery

TF. The results replicate and extend findings of the Lewis *et al.* study. TF shows consistent response decrement to stimulus repetition and response increment to stimulus change. Furthermore, visual orientation was accompanied by both behavioral and cardiac components that exhibited similar response patterns.

Cardiac. The results of the present experiment support the notion that cardiac deceleration is a component of the orienting response. The initial response to a new stimulus was cardiac deceleration. This response was observed to decrease with successive repetitions of the stimulus and recover when the stimulus was altered or replaced by a new one.

Sokolov (1963) suggests that autonomic feedback to central neural structures modifies the effects of stimulation and should, therefore, be associated with OR. Lacey and Lacey (1959, 1964) have found that HR deceleration is associated with facilitation of sensory intake and HR acceleration accompanies stimulus rejection. Graham and Clifton (1966), in their review of the relation between heart rate and the orienting response, conclude that HR deceleration satisfies the criteria usually associated with an orienting response. The present results provide further support for this notion as well as evidence for Sokolov's contention that the OR should have autonomic components.

Behavior. The behavioral patterns associated with TF response decrement and recovery might also be considered as signs of emotionality. The data indicate that smiling, pointing, and surprise decrease with stimulus repetition and are elicited by stimulus change. Thus, stimulus change or novelty is the occasion for affective change. Schacter and Singer (1962) and Pribram (1967) have discussed the importance of cognitive variables in emotion. According to Pribram, the physiological state underlying emotion or what has traditionally been called "arousal"

can be defined as "uncertainty" in the central nervous system arising from violation of cognitive expectancies. An important part of the experience of emotion is derived from the internal adjustments by which the system tries to reduce this uncertainty. Thus, violation of expectancy, a cognitive experience, induces the physiological arousal associated with the production of any emotional state. The present behavioral data are consistent with this formulation.

Response Decrement and its Meaning

The form of the response-decrement function, as determined by linear and curvilinear regression analyses from our data, support Thompson and Spencer's (1966) statement that response decrement follows a negative exponential function. While some of the data presented in this paper do fit linear functions, they consistently fit negative exponentials even better. The points at which both a linear and negative exponential would fit equally well would be the early trials in which the most rapid decrement occurs. Six trials of repeated stimulation may not be enough to determine uniquely the shape of the function. Further studies might extend these results by using longer trial sequences. However, the consistency of the form and exponents of the best fit curves over variables and experiments indicates the appropriateness of the negative exponential in describing the response-decrement process.

An important implication of Sokolov's model is that response decrement to repeated stimuli is a measure of the speed of model acquisition and that the rate of decrement indicates the efficiency of the model-building system. Since model acquisition involves central processes, such as memory or information storage, response decrement should also be related to performance on other cognitive tasks. In order to test this, performance on a series of concept-formation tasks, given as part of the longitudinal program, was compared to response decrement. In each concept-formation task, *S* was shown an array of three items of which two were identical with respect to either color, form, size, or number. *S* was to indicate which two items were "the same." The procedure was identical to that used by Lee (1965) in her test of concept formation in preschool children. The measure of response decrement for visual fixation was total number of seconds looking on trial 1 minus number of seconds looking on trial 6. This was corrected for initial fixation by dividing by the fixation score for trial 1. A mean score over all four sets was obtained. Response decrement was related to number of errors in the concept formation series such that the greater the response decrement, the fewer errors *S* made ($rho = .37$, $p < .05$). Thus, response decrement was shown to be related to a measure of cognitive capacity.

Response Recovery

A consistent finding for both experiments and for all variables was that the alteration in set B produced less response recovery than that for the other sets. Since the order of sets was the same for all *Ss*, it is not possible to analyze the data for stimulus differences. However, the consistent lack of response recovery to S_2 in set B suggests the need for further investigation of the conditions under which change produces response recovery. While discriminability of S_2 from S_1 is a necessary condition for response recovery, it is almost certainly not a sufficient condition (Lewis *et al.*, 1967). The nature of the S_1–S_2 change is highly relevant. Recent work by Bernstein (1967) and Zimny and Schwabe (1965) supports this by showing that the magnitude of a GSR orienting reflex is a function of the intensity of S_2. The investigation of the conditions affecting the magnitude of the orienting reflex is necessary in order to determine to which environmental changes the child is most responsive.

REFERENCES

Berlyne, D. E. *Conflict, arousal and curiosity*. New York: McGraw-Hill, 1960.

Bernstein, A. S. The orienting reflex and direction of stimulus change. Paper presented at the Eastern Psychological Association Meeting, Boston, April, 1967.

Cantor, G. N. Responses of infants and children to complex and novel stimulation. In L. P. Lipsitt and C. C. Spiker (Eds.), *Advances in child development and behavior*. New York: Academic Press, Vol. 1, 1963.

Graham, F. K., and Clifton, R. K. Heart rate change as a component of the orienting reponse. *Psychological Bulletin*, 1966, 65, 305–320.

Lacey, B., and Lacey, J. I. Cardiac deceleration and simple visual reaction time in a fixed foreperiod experiment. Paper presented at the Society for Psychophysiological Research, Washington, D. C., October, 1964.

Lacey, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein and M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1959.

Lee, L. C. Concept utilization in preschool children. *Child Development*, 1965, 36(1), 221–228.

Lewis, M. Exploratory studies in the development of a face schema. Paper presented at the Symposium on "The Origins of Social Behavior," American Psychological Association Meeting, Chicago, 1965.

Lewis, M., Goldberg, S., and Rausch, M. Attention distribution as a function of novelty and familiarity. *Psychonomic Science*, 1967, 7(6), 227–228.

Pribram, K. H. The new neurology and the biology of emotion: A structural approach. Paper presented at the Eastern Psychological Association Meeting, Boston, 1967.

Schacter, S., and Singer, J. E. Cognitive, social and physiological determinants of emotional state. *Psychological Review*, 1962, 69(5), 379–399.

SOKOLOV, Y. N. *Perception and the conditioned reflex.* New York: Macmillan (English translation, S. W. Waydenfeld, Trans.), 1963.

THOMPSON, R. F., AND SPENCER, W. A. Habituation: A model phenomenon for the study of neuronal substrates of behavior. *Psychological Review,* 1966, 173, 16–43.

WELFORD, N. T. The SETAR and its uses for recording physiological and behavioral data. *IRE Transactions Biomedical Electronics,* 1962, Vol. BME-9 (3), 185–189.

ZIMNY, G. H., AND SCHWABE, L. W. Stimulus change and habituation of the orienting response. *Psychophysiology,* 1965, 2, 103–115.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 81–95 (1969)

Operant Procedures for the Establishment of Stimulus Control in Two-Year-Old Infants[1]

PAUL WEISBERG[2]
Brown University

Eight 15 to 25-month-old infants were trained by multiple and chained schedules to lever press at high rates for reinforcements in the presence of S+ and at low rates when unreinforced in the presence of S−. For 4 *Ss*, S+ and S− were a steady light and 1 cycle/sec flash respectively. The stimulus assignment was reversed for the other 4 *Ss*. Usually either a VI 15-second, VI 25-second, or VR-15 schedule was programmed in all S+ components while the following schedule conditions in S− changed across *Ss* and over training sessions: extinction (EXT); the differential reinforcement of other behaviors (DRO); and a combination of the two, programmed sequentially (EXT DRO). S+, S−, and 6 flickering lights from 2 to 10 cycle/sec were presented during two generalization tests conducted in extinction. Total number of responses decreased over test sessions, and within each test, rate of responding fell progressively over trial blocks. Peak responding was always to the former S+ value. The shapes of individual gradients were much flatter across the flicker range when S− had been a 1-cycle/sec light than when it was steady.

In most operant procedures used to establish exteroceptive stimulus control, responding in the presence of one stimulus (S+) is intermittently reinforced while responding in the presence of another stimulus (S−) is never reinforced. Thus in discrimination studies with normal children (Long, 1962, 1963) and retarded individuals (Bijou and Orlando 1961; Orlando, 1965), S+ and S− were correlated with separate reinforcement schedules and each stimulus-schedule condition was presented in a suc-

[1]The major part of this research was supported by a USPHS Grant (MH 11699-01) awarded to the author. Pilot work was supported in part by USPHS Grant (NB 04268) awarded to Dr. Lewis P. Lipsitt. The author is grateful to Drs. Frederick W. Hegge and Scot Blue for their helpful criticism. Special thanks are due to Mrs. Jane E. Walsh and the staff of Lakeside Shelter, Warwick, Rhode Island, and to Mr. Walter W. Pippit of Children's Friend and Service, Providence, Rhode Island for their excellent cooperation.

[2]Now at Department of Psychology, University of Alabama, University, Alabama 35486.

cessive fashion. Sometimes the duration of each condition was governed solely by an environmental criterion (as in multiple schedules), such as a clock set to run for a fixed time period, and sometimes by a response criterion (as in chained schedules), such as no responses for 20 seconds during S− or the time taken to receive three reinforcements during S+. In these procedures, since possible influence by the individual S+ and S− is confounded with schedule order or the delivery of a fixed number of reinforcements likely to indicate the termination of an S+ period, independent methods to assess the source of control are of importance.

In general the origin and degree of stimulus control in operant studies with children has been evaluated in two ways: by stimulus reversal and by a tandem schedule (Long, 1963). In the first procedure the two exteroceptive stimuli previously associated with different schedule conditions are frequently interchanged with one another, while in the second only one stimulus, either S+ or S−, is explicitly programmed in both schedule components so that an alteration in schedule conditions is unaccompanied by an external stimulus change. Therefore, if the separate stimuli actually exercised control, both procedures would initially engender a marked disruption in the pattern of responding commonly observed prior to the imposed manipulation.

Another procedure used to identify controlling stimuli is that of generalization. Here, the training stimuli used during discrimination, together with a number of novel stimuli, each graded in terms of similarity to S+ and S−, are presented and responding in the presence of each is not reinforced. Thus in the generalization paradigm, both the training stimuli (and others similar to them) are correlated with a new schedule (extinction) and in a new order. Moreover, evidence for stimulus control during generalization testing is indicated not by a process of behavioral impairment, but by the former training stimuli continuing to produce the same relative differences in responding that existed during discrimination training.

This experiment sought to institute operant discrimination techniques in 15 to 25-month-old infants, an age group that can be profitably studied over several experimental sessions (Weisberg and Fink, 1966; Weisberg and Tragakis, 1967). To evaluate sources of behavioral control, generalization testing was done during which several stimulus values were administered under conditions of extinction. Assessment techniques utilizing multiple stimulus values rather than two points on a continuum enables one to test the limits of sensory discrimination more economically and supplies greater information about the stimulus dimension underlying variations in behavior.

METHOD

Subjects, Apparatus, and Setting

Eight infants in temporary residence at an institution were tested in an unused room at the institution. Except where noted, the *Ss* were run once a day, seven days a week, and only rarely did more than 24 hours intervene between sessions. Daily snacks usually given to all children were omitted for experimental *Ss*. Subjects were tested 2 to 3 hours after lunch. The first three columns of Table 1 identify the *Ss* in terms of the order in which they were run, along with sex and age characteristics. Prior to the first session of magazine training, the length of institutionalization for individual *Ss* varied from 9 to 207 days.

During each session, the child sat and was strapped into a high chair from which the feeding tray had been removed. Before him was a large wooden frame that housed a 70-bucket Gerbrands Universal Feeder. The entire superstructure rested on a small table. A red plastic lever extended from a chassis box that was mounted on the right front-side of the feeder frame; the lever was easily in reach of *S* and required a downward force of 26*g* through 2 mm for a response to be recorded. A Plexiglass food-

TABLE 1
DISCRIMINATION TRAINING PROCEDURE

Subject	Sex	Age at start of discrimination training (month-day)	Discrimination training[a]
S1	M	25–21	Mult VI 15 EXT(7)
S2	F	18–25	Mult VI 15 EXT(9); Mult VI 25 EXT(1); Chain DRO VI 25(2)
S3	M	18–1	Mult VI 25 EXT(6); Chain EXT DRO VI 25(8)
S4	F	16–27	Mult VI 25 EXT(3); Chain EXT DRO VI 25(8)
S5	M	15–29	Mult VI 25 EXT(3); Chain EXT DRO VI 25(13)
S6	M	15–22	Chain DRO VR 15(3); Chain EXT DRO VR 15(3)
S7	M	16–3	Chain DRO VR 15(10); Chain EXT DRO VR 15(2)
S8	M	20–10	Chain DRO VR 15(5); Chain EXT DRO VR 15(2)

[a] Numbers in parentheses denote the number of sessions devoted to the particular schedule.

collecting tray was centrally mounted to the bottom front of the feeder support. In the same room about 6 feet from the rear of the feeder was a movable cart that held programming circuitry, power supplies, electric counters, a cumulative recorder, and a device to generate different flicker rates. Each lever press produced an audible click made by the stepping motor of the cumulative recorder.

An incandescent indicator light, with an 11/16-inch unfrosted green glass dome over it, served as the light source. The light was mounted 1¾ inches above and to the left of the lever. Interruption of the light by the flicker stimulator produced flashes up to 10 cycle/sec with an error of ±5% of the specified flicker value as calibrated by a photo-sensitizing device in conjunction with an oscilloscope. Each light flash lasted 70 msec. Two observers judged the brightness of the steady light (using an SEI exposure photometer) to be 3.36 log ft.-lamberts. The flicker stimulator consisted of a 60 rpm synchronous motor that simultaneously drove a series of cams, on the surface of which protruded a different number of small nodules (varying from none to ten). Each cam struck against independent microswitches which in turn were fed into the programming circuitry. When one flicker value changed to another, no auditory cues were provided (silent push buttons were used) and all stimuli were programmed by noiseless solid-state circuitry. The cam-microswitch action did, however, cause a constant clatter in the room that conveniently masked any outside hallway noises.

One *E* was constantly in the room with the child; he sat behind the equipment cart and monitored the control unit and recorded responses. Most of *S*'s view of *E* was obscured by the proper positioning of the cart, although if *S* leaned over far enough, he could catch glimpses of *E*. Infrequently, *S* spent time in search of lost or misplaced snacks. After assisting in their recovery, *E* returned to his station behind the cart. Such very brief and rare interruptions usually came early in acquisition training and were the only social contact initiated by *E* during a testing session.

Magazine Training

The details of this technique are described elsewhere (Weisberg and Fink, 1966). Briefly, reinforcements consisting of small pieces of Cinnamon Crisp cookies (Keebler) were dispensed to *S* who sat before the feeder with the lever removed. *E* first sat alongside and encouraged *S* to eat. After the child repeatedly looked in the vicinity of the food tray whenever the feeder motor signalled the delivery of a reinforcement, *E* moved further away and faded out on social interactions. During the last few trials of magazine training, *E* remained behind the equipment cart.

Care was taken not to reinforce any apparent supersitious acts and in general no new snacks were delivered until *S* had finished eating the previous one. From 30–35 reinforcements were given within 10–15 minutes.

Rate-Strengthening

After the magazine-training session, all *S*s were reinforced with snacks for every lever press. Neither instructional stimuli nor shaping procedures were used; the *S*s were left to "discover" the CRF contingency. *S*1 and *S*2 received two sessions of CRF training (50 reinforcements per session) and were then given discrimination training.

The other *S*s were placed on intermittent schedules of reinforcement prior to discrimination training. Intermittent reinforcement was instituted, either during the initial CRF session or in the very next session, in order to strengthen and sustain a rate of 20 or more responses per minute during the last 15 minutes of two consecutive sessions. Most *S*s received intervening training on small FR schedules before some were switched to VI 25-second schedules in which the inter-reinforcement interval ranged from 5 to 40 seconds. (*S*3, *S*4, *S*5) or to VR 15 schedules in which the ratio values ranged from 5 to 40 responses (*S*6, *S*7, *S*8). Each *S* received no less than four rate-strengthening sessions during which 50 reinforcements were usually dispensed.

During the final rate-strengthening session the indicator light, which had been disconnected and nonfunctional, was turned on and responses in its presence were reinforced according to the prevailing schedule. For *S*3, *S*4, *S*6, and *S*8, the light (S+) flashed once per second and for *S*1, *S*2, *S*5, and *S*7, S+ was a steady light. S− was not presented until discrimination training, and no reinforcements were given during any S− interval. S− for the first group of subjects mentioned above was a steady light and for the other group, it was a 1-cycle/sec flash.

Discrimination Training

Table 1 indicates the kind of multiple (mult) and chained (chain) schedules or reinforcement that were used to establish stimulus control (Ferster and Skinner, 1957). Throughout part of all of the early training sessions of the first six *S*s, multiple VI 15 EXT or multiple VI 30 EXT schedules were instituted. At some point, chained contingencies were put into effect, either to replace the multiple schedule (*S*2, *S*3, *S*4, *S*5) or to act as the initial schedule condition (*S*6, *S*7, *S*8). In the case of *S*2 a schedule that differentially reinforced other behaviors (DRO) was associated with S−, replacing the previous extinction component in the mult VI 25 EXT schedule. During the DRO component a clock set to

run for a prearranged time continually reset itself whenever a lever press occurred in S−. After the clock timed out, the stimulus conditions changed from S− to S+. The purpose of the reset condition was to prevent the subject from lever pressing near the end of an S− period. Lever pressing in S+ by *S*2 continued to be intermittently reinforced on a VI 25 basis (chain DRO VI 25).

The early sessions of *S*6 and *S*8 and most of those of *S*7's also had the DRO component correlated with S−, but responses in S+ were reinforced according to a VR 15, thus constituting a chain DRO VR 15. During the latter part of training of the last six *S*s, two conditions programmed in sequential order defined the S− period. In the initial two-thirds, lever presses were subject simply to extinction and in the final one-third, the DRO contingency was in effect. The chain arbitrarily ended in S+ when, depending upon *S*, (*cf.* Table 1) the VI 25 seconds or VR 15 was in force.

In regard to the temporal properties of the discriminative stimuli, S+ durations early in training were usually three times longer than those of S− but this discrepancy was gradually reduced over training sessions until S+ intervals became either equal to S− or no more than twice its size. The terminal S+ durations for *S*2 formed an exception to this rule, remaining relatively long (from 1 to 2½ minutes per exposure) from the beginning of training. In the multiple schedules of *S*2 as well as the multiple and chained schedules for *S*2, *S*3, *S*4, *S*5, S+ presentations stabilized at about 1 minute each. Comparable values for S− periods showed greater variations: in the mult VI 15-second EXT sessions of *S*1 and *S*2 the modal S− period during extinction was 50 seconds (range 25 to 60 seconds) while in the mult VI 25-second EXT sessions of *S*3, *S*4, *S*5, it was 30 seconds (range 30 to 40 seconds).

Finally, for *S*6 and *S*7, the length of S+ periods was kept consistently at 30 seconds after the first few sessions. In the case of *S*8, however, individual S+ periods were not fixed by a clock; determined instead by how long it took to accumulate all the available reinforcements that happened to be programmed under the existing VR 15. A mode of three was available (range 1 to 4) that amounted to an average S+ duration of 30 seconds.

Discrimination sessions during the final stages of training were 25 to 30 minutes long, during which time S+ and S− were each presented about 20 times and from 50 to 60 reinforcements were received.

Generalization Testing

Two sessions were devoted to these tests. A 10-minute period of discrimination training preceded each one, and an entire session of training also intervened between the two sessions. *S*1, however, received no inter-

polated practice. Throughout the preliminary period and the intervening session, the same stimulus and reinforcement conditions prevailed that were described for the last pregeneralization session (*cf.* Table 1).

Generalization tests were conducted under extinction conditions. Eight stimuli, including the two training stimuli, were presented, each one lasting for 30 seconds. The eight stimuli were randomly arranged within each of six blocks, making a total of 48 presentations. A 3-second "black-out" in which the stimulus source was darkened separated trials.

RESULTS AND DISCUSSION

Discrimination

Efficacy of discriminative behavior for the first and second halves of each session were assessed by the computation of Discrimination Indexes (Sadowsky, 1966):

$$\text{D.I.} = \frac{\text{total } S+ \text{ responses}}{\text{total } S+ \text{ responses} + k(\text{total } S- \text{ responses})}$$

where $k = S+/S-$ duration and was used to weight the disproportionate amount of time usually devoted to S−. The lower and upper limits of the D. I. are .00 and 1.00 respectively, with .50 representing equal response probability to S+ and S−. Table 2 gives D.I. values for the last two pregeneralization sessions of each subject as well as the session that intervened between generalization tests.

The D.I.'s were substantially higher in the single postgeneralization session compared to the last two pregeneralization sessions with most sub-

TABLE 2
DISCRIMINATION INDEXES

	Pregeneralization				Postgeneralization	
	Next to last session		Last session		Intervening session	
Subjects	First half	Second half	First half	Second half	First half	Second half
S1	.82	.71	.82	.92	—	—
S2	.71	.71	.95	.76	.94	.89
S3	.62	.89	.67	.88	.76	.74
S4	.59	.89	.65	.88	.94	.89
S5	.73	.88	.80	.80	.93	.89
S6	.75	.67	.79	.87	.77	.89
S7	.64	.76	.72	.84	.92	.80
S8	.69	.62	.91	1.00	.99	.97
Median	.700	.735	.795	.875	.930	.890

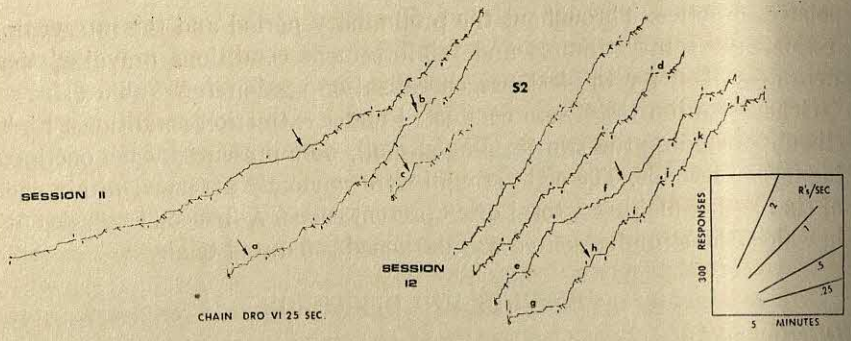


FIG. 1. Cumulative records of the last two pregeneralization sessions for *S2*. The prevailing schedule is chain DRO VI 25 seconds. Vertical marks below the curve indicate simultaneous onset of S+ and offset of S—. Vertical marks above the curve indicate simultaneous offset of S+ and onset of S—.

jects also improving in discriminative ability from the next-to-last to the last pregeneralization session. Considering D.I. changes within the pregeneralization sessions, most *S*s improved in performance with practice but this trend was slightly reversed for 5 of 6 *S*s in the postgeneralization session. In the discrimination sessions that preceded the next-to-last one in Table 2, the majority of D.I.'s were centered around .50 (range .40 to .80).

The cumulative records in Figs. 1, 2, and 3 show the rates of responding to successive presentations of S+ and S— under different chained schedules of reinforcement. Not shown are any records from sessions in which multiple schedules were used. Although the multiple VI 15-second EXT schedule during *S1*'s last session generated fairly good behavioral control (response rates in S+ were nearly five times greater than in S—), the high density of reinforcement during the VI 15-second component caused a low S+ response output. Resort to multiple VI 25-second EXT schedules with the next four *S*s elevated S+ responding greatly but S— responding also increased and there was little evidence of differential responding. To circumvent the possibility that the onset of S+ had become a conditioned reinforcer during multiple schedule programming, the DRO procedure was installed.

Figure 1 reproduces the performance of *S2*'s last two chain DRO VI 25-second sessions. In contrast to intensive S+ responding, the majority of S— periods illustrate a low level of lever-pressing activity but there are several places (arrows) where S— responding is mixed unevenly between high rates and periods of nonresponding that fell short of the DRO interval. This erratic response pattern occurred mostly during periods *a* to *l* inclusive when the DRO interval was 30 seconds; in all other S—

periods, it was 20 seconds or lower. *S2*'s difficulty to completely suppress behavior during S− periods having DRO 30-second requirements gave rise to the alternative and final strategy, used with the remaining *S*s, of keeping the DRO interval short (10 seconds) and to precede it with a longer extinction component (20 seconds). The duration of S− periods was kept long (30 seconds) for two reasons: as preparation for the upcoming generalization trials each of which was 30 seconds long; and to help sharpen the contrast during discrimination training between periods of unreinforced as against reinforced responding.

Figures 2 and 3 respectively reveal terminal discrimination performance when the EXT DRO component was combined in sequence either with the VI 25-second schedule (*S3*, *S4*, *S5*,) or with the VR 15 schedule

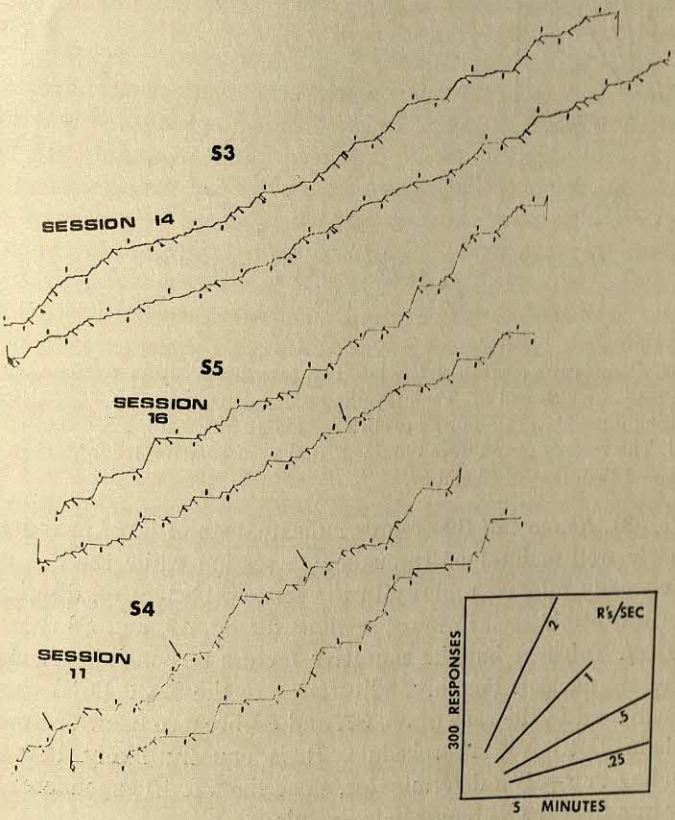


FIG. 2. Cumulative records of the last pregeneralization session for *S3*, *S4*, and *S5*. The prevailing schedule is chain EXT DRO VI 25 seconds. Vertical marks below the curve indicate simultaneous onset of S+ and offset of S−. Vertical marks above the curve indicate simultaneous offset of S+ and onset of S−.

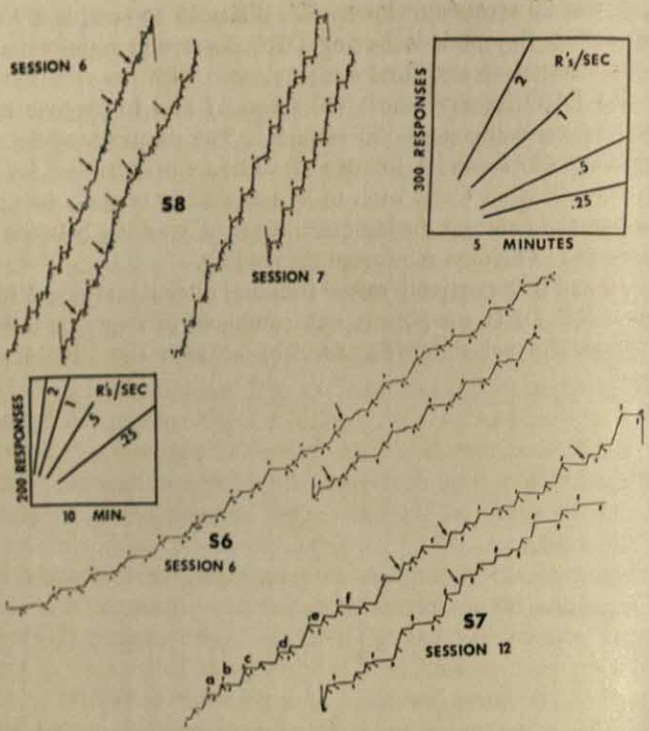


FIG. 3. Cumulative records of the last two pregeneralization sessions for *S*8 and the last session for *S*6 and *S*7. The prevailing schedule is chain EXT DRO VR 15 except during intervals *a* to *e* in *S*7's twelfth session at which the EXT component was omitted. The records for *S*8 were obtained from a cumulative recorder that ran at a slower speed than that of *S*6 and *S*7.

(*S*6, *S*7, *S*8). As most of the records indicate, rate of lever pressing in S+ was fairly well maintained throughout a session while the S− rate became virtually nil once midway in the session. Generally, discriminative control was established in fewer sessions for *S*6, *S*7, and *S*8 than for *S*3, *S*4, *S*5 (*cf.* Table 1) but the causative factors responsible for these differences are unclear. Possibly, adjustment to the EXT DRO component was facilitated in the case of *S*6, *S*7, and *S*8 because of prior exposure to the chain DRO VR 15 schedule. It is equally likely that rate of responding in S+ is a determinator, since the VR 15 engendered a much higher rate than the VI 30-second schedule.

The last two records of *S*8 (Fig. 3) are provided to demonstrate a case of sudden transition in differential responding. As is evident, the uniformally high S− rates characteristic of session 6 are sharply attentuated

in session 7, when responding to the last ten presentations of S− resulted in "errorless" performance. Acquisition of behavioral control for the other *Ss* was never achieved so rapidly as that of *S8*.

It is probable that had the DRO component been eliminated from a chain that had already produced reliable differences in lever pressing to S+ and S−, the degree of established control would have remained intact. This claim is partially supported by the stable transfer of behavioral control from the discrimination procedure to the generalization test series (*cf.* Fig. 5) that employed an extinction procedure. Bijou and Orlando (1961) and Orlando (1965) reported no disruption in differential responding by retardates to alternate presentations of S+ (correlated with a variable ratio schedule) and S− (correlated with extinction) after successful discrimination performance with a chained schedule in which the S− stimulus was associated with a DRO component. In addition, Long's (1962) discrimination data with children "suggest that if strong stimulus control has once developed, the delay contingency (DRO component) can be removed without serious or permanent loss of stimulus control" (p. 446).

The results of discrimination training clearly demonstrate that the various schedules of reinforcement can exert a strong influence over the behavior of young preverbal children. The approximately 50 snacks presented during each session in this experiment sufficed to maintain behavior and there was no evidence of a response decrement due to food satiation. This encouraging motivating feature together with the rare occurrence of recalcitrant behavior by the toddlers bodes well for the use of children of this age in further operant studies.

Generalization

Weighted D.I.'s based upon S+ and S− presentations during the ten-minute discrimination periods immediately preceding testing, were all above .85 with the following exceptions: *S2*, second test session (D.I. = .82); *S1*, first test session (D.I. = .78), and *S6*, first test session (D.I. = .82).

Figure 4 shows that over the course of trials in both sessions of generalization testing, the total number of responses progressively fell for most subjects. This response decrement, an extinction effect, supports a common finding (Guttman and Kalish, 1956; Jenkins and Harrison, 1960) and where reversals in the curves appear (as in the low output during the second sessions of *S1* and *S3* in the first session of *S8*), they were caused almost entirely by the subjects showing a renewed surge of responding when the former S+ stimulus was reintroduced. As a second trend, most subjects responded more throughout the first generalization session than

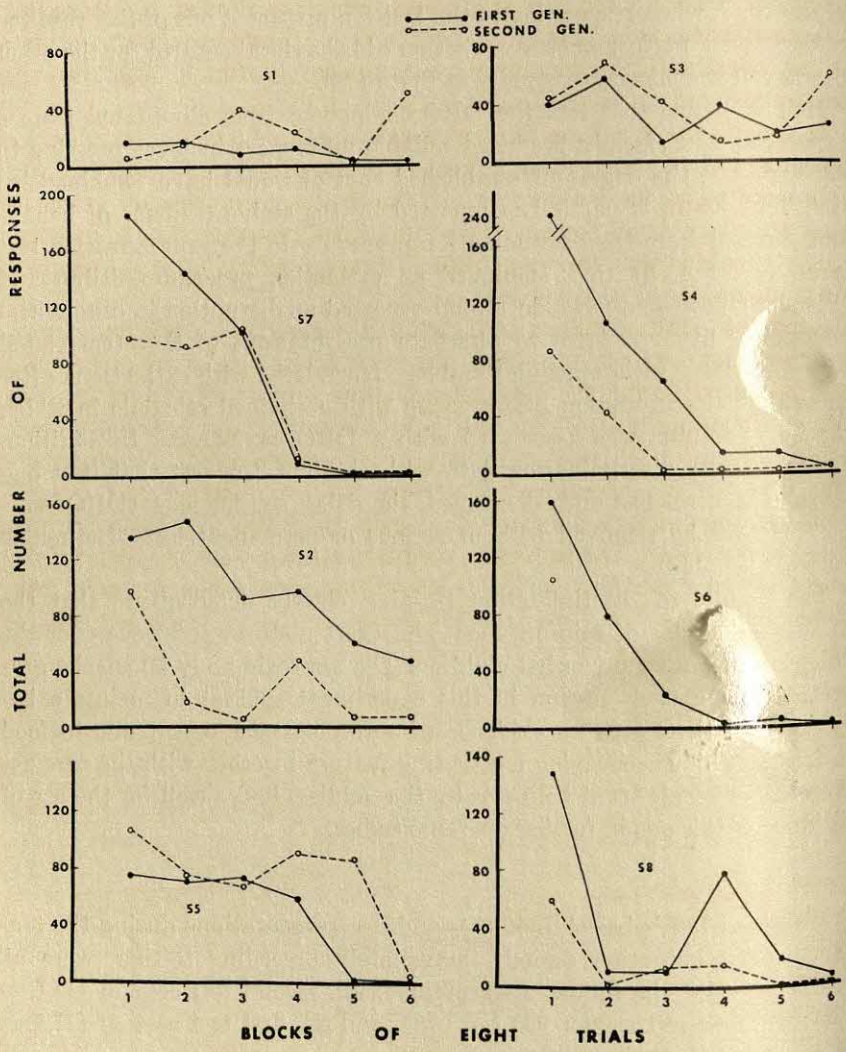


FIG. 4. Changes in total number of responses over trial blocks for the first and second test sessions.

the second (Jenkins and Harrison, 1960; Newman and Baron, 1965). Again the extinction records of *S*1 and *S*3 pose exceptions, with the between-session response difference in favor of the second test day, although this discrepancy is not consistent across trial blocks. The extinction function for *S*6's second test session is incomplete because he started to cry midway during the second trial block.

Responses to each of the eight test stimuli, ordered according to flicker

frequency, were pooled over trials, and proportional rates of responding to each stimulus were determined by the ratio: total number of responses to test stimulus/total number of responses to all stimuli. The resulting relative generalization gradients are plotted in Fig. 5.

In general the gradients for both sessions parallel each other, with the percentage of responses always greatest to the previous S+ stimulus and even more so on the second than the first test session. Sharper gradients after an interpolated session of discrimination training confirm previous results (Jenkins and Harrison, 1960; Terrace, 1966b); after no practice, the gradients for *S*1 also steepen.

Since the relative difference between S+ and S− responding during generalization testing was within the same limits and often better than that produced during the last discrimination session (*cf.* Table 2), with the confounding effects of the schedule order and reinforcement-nonreinforcement delivery ruled out during generalization, it would appear that the *S*s discriminated between the two exteroceptive stimuli. However, the stimulus dimensions(s) upon which this discrimination was based is not clear. The presently used test stimuli were not equated for brightness level and this quality is known to increase monotonically with flicker frequency (Geldard, 1953). Moreover, the 10-cycle/sec light is

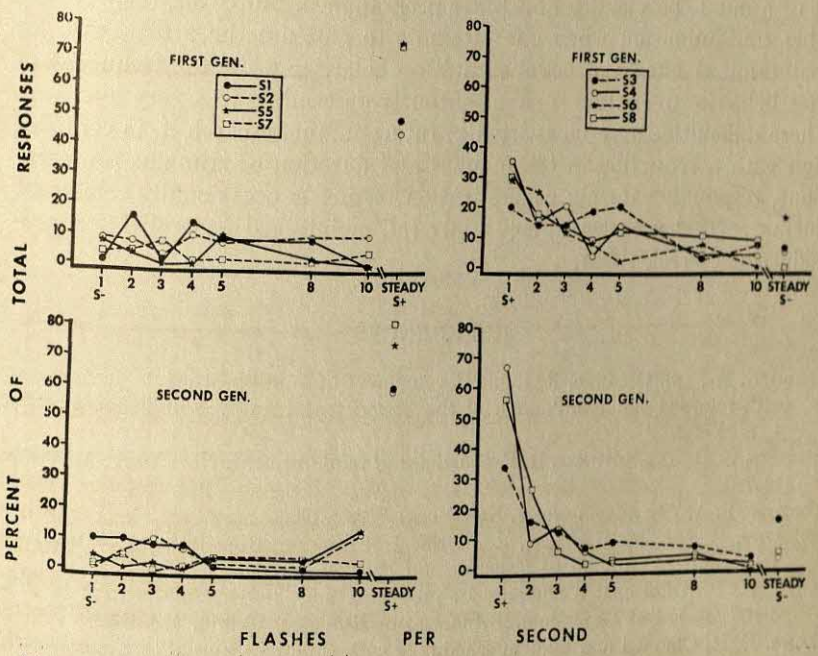


FIG. 5. Relative generalization gradients for the first and second test sessions.

probably well below the critical fusion frequency (Misiak, 1951). Hence, the gradients on the left of Fig. 5 may represent response to a series of several flickering lights and a steady one or a series of several dim lights and a bright one.

The decremental gradients on the right of Fig. 5 suggest that some *Ss* may have been attending more precisely to each stimulus value instead of responding exclusively on a "go" - "no go" basis. For example, on the second generalization test session, *S*8's behavior very accurately ordered the physical continuum, with a relatively large proportion of responses emitted to the 2 cycle/sec light. Here, too, as with the other *Ss*, the controlling stimulus continuum could be one of intensity or flicker frequency; had additional generalization testing been performed with a set of steady lights, each different in brightness level but representative of the levels created by the presently employed flicker set, a more positive position perhaps could have been taken (*cf.* Terrace, 1967a).

Finally, a correct interpretation of generalization data gathered by extinction procedures poses problems. Emotional effects due to removal of reinforcements for a prolonged period may result in a loss of control: response bursts may occur in the presence of S− or stimuli proximal to it and/or unduly long pauses in responding may persist even when stimuli at or around the vicinity of S+ are programmed. Moreover, there is possible contamination when one attempts to measure the attenuative (or maintaining) effects of novel stimuli on behavior while the frequency of that behavior over time is simultaneously descending to a very low level. These difficulties may be averted in future infant research if, in conjunction with a reduction in the number and duration of stimulus presentations, responding during generalization testing is occasionally reinforced during several preselected S+ trials (*cf.* maintained generalization procedure of Pierrel, 1958).

REFERENCES

Bijou, S. W., and Orlando, R. Rapid development of multiple-schedule performance with retarded children. *Journal of the Experimental Analysis of Behavior*, 1961, 4, 7–16.

Ferster, C. B., and Skinner, B. F. *Schedules of reinforcement*. New York: Appleton, 1957.

Geldard, F. A. *The human senses*. New York: Wiley, 1953.

Jenkins, H. M., and Harrison, R. H. Effect of discrimination training on auditory generalization. *Journal of Experimental Psychology*, 1960, 59, 246–253.

Long, E. R. Additional techniques for producing multiple-schedule control in children. *Journal of Experimental Analyses of Behavior*, 1962, 5, 443–455.

Long, E. R. Chained and tandem scheduling with children. *Journal of Experimental Analysis of Behavior*, 1963, 6, 459–472.

MISIAK, H. The decrease of critical flicker frequency with age. *Science*, 1951, **113**, 551–552.

ORLANDO, R. Shaping multiple schedule performances in retardates: establishment of baseline by systematic and special procedures. *Journal of Experimental Child Psychology*, 1965, **2**, 135–153.

PIERREL, R. A generalization gradient for auditory intensity in the rat. *Journal of Experimental Analysis of Behavior*, 1958, **4**, 303–313.

SADOWSKY, S. Discrimination learning as a function of stimulus location along an auditory intensity continuum. *Journal of Experimental Analysis of Behavior*, 1966, **9**, 219–225.

TERRACE, H. S. Discrimination learning and inhibition. *Science*, 1966(a), **154**, 1677–1680.

TERRACE, H. S. Stimulus control. In W. K. Honig (Ed.), *Operant behavior: Research and Application*. New York: Appleton, 1966(b).

WEISBERG, P., AND FINK, E. Fixed ratio and extinction performance of infants in their second year of life. *Journal of Experimental Analysis of Behavior*, 1966, **9**, 105–109.

WEISBERG, P., AND TRAGAKIS, C. J. Analyses of DRL behavior in young children. *Psychological Reports*, 1967, **21**, 709–715.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 96–100 (1969)

Reinstatement of an Operant Response by the Delivery of Reinforcement during Extinction[1]

JOSEPH E. SPRADLIN, DEAN L. FIXSEN,[2] AND FREDERIC L. GIRARBEAU[3]
Bureau of Child Research, University of Kansas, and Parsons State Hospital and Training Center

Twelve severely retarded children were conditioned to respond at stable rates for 4 consecutive days on a fixed-ratio schedule. They were then given a single extinction session. The extinction session lasted until they had met a 2-minute pause criterion 15 times. On five of the occasions after the *Ss* met the pause criterion, they were delivered reinforcement; on five other occasions a buzzer sounded; and on another five occasions no change in the environment occurred (control). The *Ss* made more responses after reinforcement than after the buzzer or after the control.

Recently, Spradlin, Girardeau, and Hom (1966) compared the number of responses following noncontingent (free) delivery of reinforcement during extinction with the number of responses following a control period during which there was no stimulus change. The results clearly indicated that the free delivery of reinforcement in extinction reinstated behavior which had previously been maintained by contingent reinforcement. The present study was designed to replicate the reinstatement effect with nonverbal children using a nutrient reinforcer with an added control to determine if reinstatement could be established by the presentation of an abrupt novel stimulus.

METHOD

The subjects were 12 severely retarded children between 7 years, 4 months and 14 years, 11 months of age. None of the subjects could be tested with such standardized tests as the Revised Stanford Binet or the Wechsler Intelligence Scale for Children. Their social age equivalents on the Vineland ranged from 1 year, 4 months to 3 years. Their social quotients ranged from 14 to 40. None was classified by the aide in charge of

[1] This study was supported by NICHHD grant HD 00870.

[2] Dean L. Fixsen is now a predoctoral trainee with the Bureau of Child Research, University of Kansas, Lawrence, Kansas.

[3] Frederic L. Girardeau is now Director of Research Training, Bureau of Child Research, University of Kansas Medical Center, Kansas City, Kansas.

his cottage as having functional speech. There were seven males and five females. The subjects had to meet the following criteria to be included in the reinstatement study: They had to eat M&Ms, cereal, mints, or marshmallows when they were offered to them; they had to be able to grasp and hold an object (pencil); they had to meet a stability criterion. The stability criterion consisted of four consecutive sessions of responding in which the response rate for any single session deviated from the mean of four sessions by no more than 20%. Four of the original 17 subjects, selected on the basis of the first two criteria, were eliminated because they did not reach the stability criterion. A fifth subject was discarded because of an error in experimental programing.

Apparatus and Experimental Room

The experimental room in which the *S* was placed contained a white formica-covered response panel with two recessed Grason-Stadler push-button response keys located approximately 64 inches apart with a reinforcement receptacle located midway between the two keys. Both keys were located approximately 32 inches above the floor. Below the response keys were two Lindsley manipulanda. For this experiment, both Lindsley manipulanda and the left response key were inoperable. The right response key was illuminated by a red light at all times, except when reinforcement was delivered. When reinforcement was delivered, the light behind the response key was extinguished, the house light was dimmed, and the receptacle lighted for 5 seconds. The novel stimulus was a 90 dB (SPL) complex noise produced by a Line Electric Company six-volt buzzer. Reinforcement programing and response recording was done automatically by standard programing and recording equipment housed in a room adjacent to the observation room. The distance of the programing and recording equipment from the experimental room, plus 80 dB SPL of white noise, eliminated any control which relay clicks might exert over the *S*s' behavior.

Establishing Responding on the FR Schedule

Each child was trained through demonstration and shaping procedures to press the response key. Once the child was responding, the experimenter left the room and gradually increased the ratio until an FR-25 was reached. The rapidity of the increase in ratio varied between *S*s according to their response rate. The *S* was then maintained on an FR-25 schedule until the stability criterion was reached. There were two exceptions to this procedure—S_1 and S_2 were maintained on an FR-50 schedule since both of these *S*s were high-rate responders who showed satiation effects toward

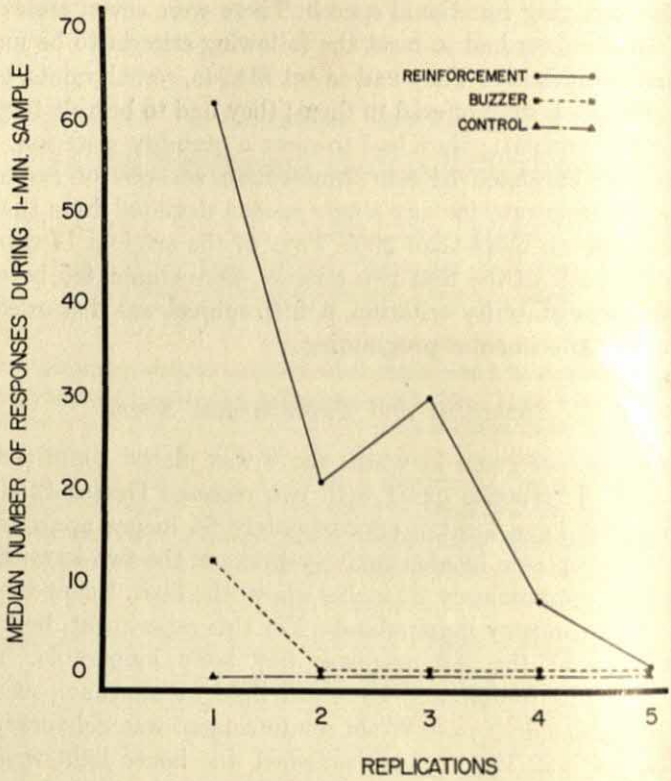


FIG. 1. Median responses during 1-minute period after the occurrence of reinforcement, the buzzer, or during the 1-minute control period.

the end of the session. During conditioning, all sessions were 15 minutes long.

Extinction of the Response and Evaluation of the Reinstatement Properties of Reinforcement Versus a Novel Stimulus

Once the *S*s met stability criterion, they were given a single extinction session. During the extinction session, the equipment was programed so that a 2-minute pause in responding resulted in the occurrence of reinforcement (plus correlated stimuli), a novel stimulus (5-second buzzer), or a control period (no stimulus change was presented). A 1-minute sample of responding was recorded immediately after the *S* met each pause criterion. The presentation of reinforcement, novel stimulus, and control period was replicated five times during the extinction period. The sequence of presentation of the three events was the same during all five

replications for a given *S*. This resulted in six possible sequences of presentations of the three events. Two *S*s received each sequence—thus completely counterbalancing sequence effects.

RESULTS

Eight of twelve *S*s responded most during the 1-minute period following reinforcement; two *S*s responded most after the buzzer; and two *S*s responded most during the 1-minute control period. A Friedmann analysis of variance with ranked data was significant at the .05 level (Walker and Lev, 1953). When pairs of treatments were compared with the signed rank test for paired observations, the number of responses made during the 1-minute period following reinforcement was significantly greater than the number of responses made after either the buzzer or control. The number of responses made after the buzzer and during the control period did not differ significantly from each other. Figure 1 shows the median number of responses made during each of the five 1-minute periods for the three conditions. A Friedmann analysis of ranks for replications was also significant at the .05 level. This finding plus inspection of the data indicate that number of responses decreased with additional replications.

DISCUSSION

The present study, as well as previous studies, demonstrate that free reinforcement delivered during extinction results in the reinstatement of the acquired response above the level observed following either a control period or an abrupt auditory stimulus change (Spradlin *et al.*, 1966; Campbell, Phillips, Fixsen, and Crumbaugh, 1968). These data lend further support to Jenkin's (1965) contention that reinforcement can develop discriminative properties during free operant training. According to this interpretation, during experimental training, a chain develops in which the *S* runs off a series of responses followed by the delivery of reinforcement. The *S* then stops responding, retrieves the reinforcement, and returns to responding again. After a series of such events, it is assumed that chaining occurs with the reinforcement serving as the S^D for resuming responding. Since reinforcement is followed only once by responding without reinforcement (onset of extinction), it is assumed that the S^D properties of reinforcement are only partially extinguished during traditional extinction. Thus, when free reinforcement is introduced during extinction, it immediately reinstates responding. The decrease in the reinstatement effect as a function of repeated presentations offers further support for this explanation.

SUMMARY

This study investigated the reinstatement of responding during extinction as a function of reinforcement versus a novel stimulus. The *Ss* used were nonverbal, severly retarded children. The *Ss* made significantly more responses during a 1-minute period following reinforcement than during either a 1-minute period following a novel stimulus or a 1-minute control period. Number of responses made after the novel stimulus and during the 1-minute control did not differ significantly.

REFERENCES

Campbell, P. E., Phillips, E. L., Fixsen, D. L., and Crumbaugh, C. Free operant response reinstatement during extinction and time-contingent (DRO) reward. *Psychological Reports*, 1968, **22**, 563–569.

Jenkins, H. M. Measurement of stimulus control during discriminative operant conditioning. *Psychological Bulletin*, 1965, **64**, 365–376.

Spradlin, J. E., Girardeau, F. L., and Hom, G. L. Stimulus properties of reinforcement during extinction of a free operant response. *Journal of Experimental Child Psychology*, 1966, **4**, 369–380.

Walker, Helen M., and Lev, J. *Statistical inference*. New York: Holt, 1953.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 101–113 (1969)

Infant Social Development: Some Experimental Analyses of an Infant-Mother Interaction during the First Year of Life[1]

ROBERT G. WAHLER
The University of Tennessee

An infant-mother interaction was studied over a period of about 1 year. During this time, experimental analyses of the infant's developing verbal behavior were conducted. Results indicated that his mother served as an effective source of reinforcement for several infant responses classes. In addition, the data revealed that some infant response classes were functionally related to older infant response classes.

Although infancy has long been considered as a formative period of great importance for the development of social behavior, only recently have researchers become interested in detecting variables that are operative during this stage. Prior to the last decade most investigators adopted normative approaches to the problem, primarily attempting to establish age limits within which certain social behaviors can be expected to appear. The last decade, however, has seen an awakening interest in the use of experimental techniques, aimed at providing cause-effect analyses of infant development. Many researchers, a few working within the conceptual framework of reinforcement theory (e.g., Gewirtz, 1956, 1961), and others postulating the existence of need systems (e.g., Bowlby, 1958, 1960), have begun to shed light on the processes involved in early social development.

Most of the experimental studies of early social behavior have been conducted by the "need system" proponents, such as Caldwell (1962), Schaffer (1959, 1963, 1964), Spitz (1945, 1946), and Wolff (1959, 1961). Reinforcement theory, while it has stimulated a great deal of research on the social development of children beyond the infancy stage (e.g., Cairns, 1961; Stevenson, 1963; Hartup, 1964), has produced few experimental analyses of this phenomenon in infants. The paucity of infancy studies using the reinforcement model is surprising, both in terms of the

[1]This study was supported in part by Research Grant MH 13703-01 from the National Institute of Mental Health. The author is grateful to Norbert Reese and John Silk, who served as observers in the study.

utility of the model in conceptualizing social development in older children and the fact that those few infancy studies that have used the model have achieved positive results. Concerning this latter point, evidence has been presented to show that such infant behaviors as smiling (Brackbill, 1958; Wahler, 1967), babbling (Rhiengold, Gewirtz, and Ross, 1961; Weisberg, 1963) and crying (Etzel and Gewirtz, 1967) may be under adult reinforcement control. It would thus seem that the reinforcement model could be fruitfully applied in more extensive analyses of early social behavior.

"More extensive analyses," for one thing, implies the fact that all currently reported reinforcement studies of infant behavior have been aimed at variables responsible for the *maintenance* of such behaviors. That is, the major goal of each study involved testing the reinforcement value of adult social attention for already present behaviors. While information is therefore available concerning reinforcement variables in the infant's current environment, little is known about his past environment. Quite possibly, social reinforcement variables are also important in the development of behavior, but this conclusion cannot be directly supported by existing data. However, longitudinal studies of individual infants could shed light on this problem area. These studies might be addressed to the following points:

If observations of an infant-mother interaction are begun shortly after birth, undoubtedly the observational records will eventually show that some class of infant behavior is frequently correlated with certain behaviors of his mother. At that point one might conduct the usually reported experimental analyses to assess the mother's reinforcement control of this class (such control of infant smiling has been reported as early as 2 to 3 months of age—e.g., Brackbill, 1958). At some further point in the infant's life, continued observation will reveal the presence of new response classes. If experimental analyses indicate that the mother also has reinforcement control of the new class, an ontogenetic question is posed: Is the new class functionally related to any of the older classes? It is reasonable to assume from the reinforcement model that the development of a new response class occurs at the cost of the older classes. That is, a new response class may be thought of as more effective means of obtaining reinforcement—more effective than a functionally similar older class. One would thus expect the older class to gradually diminish in strength, through extinction and because it may be incompatible with the new class. Eventually, then, a response strength hierarchy is formed, in which the newer class becomes the more likely response in a given setting.

If specific response strength hierarchies are formed in the infant's

development, systematic observation and experimental analyses will be helpful in demonstrating the phenomenon. When a new response class appears in the infant's behavioral repertoire, observational data should reveal predictable changes in one of the older response classes. Specifically, as the new class gains in strength, the strength of one of the older classes should gradually diminish—suggesting the beginning formation of a response hierarchy. Next, when the new class is subjected to experimental analysis, the previously identified older class should be predictably affected. If both classes are influenced by the mother's social attention, and if they exist in the same response strength hierarchy, then lowering the response probability of one of the classes should increase the response probability of the other class. Therefore, if the social contingencies of the new class are eliminated (through instructions to the mother) one would predict: (1) the strength of the new class would decline, and, (2) the strength of the older class would increase (while other response classes remain unaffected). The present study was designed to yield data relevant to these possibilities.

METHOD

Subjects, Apparatus, and Observers

The subject (*S*) was a healthy, full-term male infant who was 3 weeks of age at the beginning of the study. The *S* was obtained from a public listing of area births through the following procedure: all parents on the list were contacted through a letter requesting their participation in the study. Of 25 parents contacted, two announced their willingness to participate in the study. Both *S*s were studied initially; but because of time and scheduling problems, one of the *S*s was soon dropped from the study.

The apparatus consisted of an Esterline-Angus multichannel event recorder and two operating panels containing push-button microswitches. Depression of the microswitches by observers activated selected channels of the event recorder. This apparatus permitted the simultaneous recording of up to 20 response and stimulus classes.

Observers were graduate students in Clinical Psychology; all were sophisticated in operant techniques and natural science principles of observation. While they were aware of the general purpose of the study, they were ignorant of the sequence of procedural steps involved.

Dependent and Independent Variables

In order to evaluate the developmental hypotheses outlined earlier, it was deemed important to choose as dependent measures, infant behaviors

that would be likely to display relatively rapid and frequent changes over time. Casual observation and normative research (e.g., Shirley, 1933) suggested that verbal behaviors would meet these criteria nicely. It was therefore decided to focus exclusively on the *S*'s verbal responses.

Prior research on infant social behavior (e.g., Rhiengold *et al.*, 1959) has shown that a variety of adult social behaviors may serve as reinforcers. Physical contact with the infant, talking to him, and smiling were all effective in controlling his behavior. Apparently, much of what an adult normally does or says in social interactions can be reinforcing to the infant. It was thus decided to consider the above general class of adult behaviors as possible reinforcers for *S*. This stimulus class was referred to as "mother social attention."

Procedure

The study was conducted in the *S*'s home during the morning hours, on a regular once-a-week basis. Each session was 25 minutes in length.

At the beginning of each session, *S* was placed face up in his crib by his mother and all distracting objects removed from his range of vision. Two observers then positioned themselves at opposite ends of the crib, allowing themselves an unobstructed view of *S*. The observers remained as immobile as possible in these positions while the experimenter (*E*) gave instructions to the mother. When the session began, the mother was positioned between the observers, leaning over the crib approximately 3 feet from *S*'s face. *E* stood behind the mother in partial view of *S*.

For the first 5 minutes of the session, instructions to the mother were to play with *S* any way she wished as long as she did not pick him up. During this 5-minute segment, *E* and the observers kept a running written record of *S*'s verbal behavior. At the end of this time period, *E* and the observers discussed their records while the mother continued to play with *S*. Discussion was aimed at the classification of *S*'s verbal behavior into categories, based on topographic similarities among his separate responses (e.g., glottal sounds, including aspirate *h* or stops and catches made in the throat).

When response classes were established by the face validity grouping described above, efforts were then made to assess their frequencies during the next 20 minutes. In addition, mother social attention was recorded when contingent upon any of the classes. Instructions to the mother now depended upon whether the session was designated as *baseline* or *experimental* (see later section). The observers now made their recordings via the multichannel event recorder. A timing device was also introduced that produced an audible click every 10 seconds. The observers were instructed to press and release selected microswitches upon hearing the

click if members of any of the previously defined response classes or contingent social attention occurred during the preceding 10-second interval. At the end of each session observer agreement was computed for each response and stimulus class. To assess the reliability, an agreement or disagreement was tallied for every 10-second interval, and the percentage of agreements between observers was computed. Over all sessions, reliability checks showed observer agreement of 90% or better.

Baseline sessions. When a response class was formulated, its frequency was recorded from that point on. Baseline sessions served to assess the strength and stability of classes and to assess frequency relationships between classes. *E* examined the frequency records after each baseline session with reference to the following points: (1) frequency measures of new response classes were examined in connection with older classes; (2) special note was taken of reciprocal frequency relationships between new and old response classes. That is, it was predicted that as a new class gained in frequency over sessions, an older class would decline in frequency. When such a relationship appeared, *E* scheduled *experimental* sessions to evaluate the relationship and to assess mother reinforcement control of the new class. During the baseline sessions, the mother was instructed to be especially attentive to the new and old classes.

Experimental sessions. Experimental sessions were scheduled without informing the observers. The mother was contacted prior to the session or while the observers were preparing their apparatus at the beginning of the session. In all instances she was instructed to ignore all occurrences of the new class and told that she should "freeze" (i.e., do nothing) when response members of the class occurred; when the response in question terminated she could again be attentive in any way she pleased. *E* further informed the mother that he would signal her to "freeze" by tapping her on the shoulder. As she gained practice in the technique, *E* scheduled his signals to be less frequent.

On an arbitrary basis, *E* decided to limit the above modifications in mother social attention to three sessons. At the end of that time *E* again instructed the mother to be attentive to the new class—and to continue her attention to the older class. These instructions were in effect for the duration of the study.

If the mother was in reinforcement control of the new class and if the new class was functionally related to the older class, the following results would be predicted: (1) frequency measures of the new class should decrease and increase correlated with the absence and presence of the mother's social attention contingencies; (2) the reciprocal frequency relationships between the new and the old classes should continue to be evident over the experimental sessions.

RESULTS

The following data do not include mother social attention contingencies. Records showed that her social attention was provided continuously for all of *S*'s response classes except upon instructions by *E* to eliminate attention for specific classes. In all instances, she was completely successful in following these instructions. Therefore, rather than to needlessly complicate the following figures, mother contingency data have been excluded.

Figure 1 describes frequency measures of *S*'s Glottal Sounds and Babbling-Cooing for the third through the thirty-first week of his life. Reference to the baseline portion of this figure provides justification for considering these two classes together. Glottal Sounds appeared first in *S*'s verbal repertoire, followed 2 weeks later by Babbling-Cooing; as Babbling-Cooing gained in strength over the baseline sessions, Glottal Sounds diminished.

As predicted, eliminating mother's social attention for Babbling-Cooing was followed by reduced output of this class during the first experimental modification period. This finding, of course, lends support to the contention that the mother was an effective source of social reinforcement for *S* at least by the time he was 2 to 3 months of age. Notice also that the reciprocal relationship between Babbling-Cooing and Glottal Sounds continued during the first experimental modification period; as Babbling-Cooing was reduced in frequency, Glottal Sounds gained.

Reference to the second experimental modification period provides further evidence for mother reinforcement control and for the reciprocal relationship between Babbling-Cooing and Glottal Sounds. Correlated with the mother's resumed attention to Babbling-Cooing, this class gained in strength and Glottal Sounds decreased. Thus, it would seem reasonable to conclude that the mother's social attention was at least partly instrumental in the maintenance of Babbling-Cooing, and to conclude that this class existed in a response strength hierarchy with Glottal Sounds.

During the fourteenth and sixteenth weeks of *S*'s life, two new response classes appeared in his verbal repertoire. These classes, labeled Squealing and Coughing-Gasping are seen in Fig. 2. The classes were considered together because of their frequency relationship over the baseline sessions; as the later appearing Coughing-Gasping increased in frequency, Squealing began to decline.

Once again, mother reinforcement control of the new class (Coughing-Gasping) was demonstrated; reference to the two experimental periods shown in Fig. 2 reveals that frequency measures of Coughing-Gasping

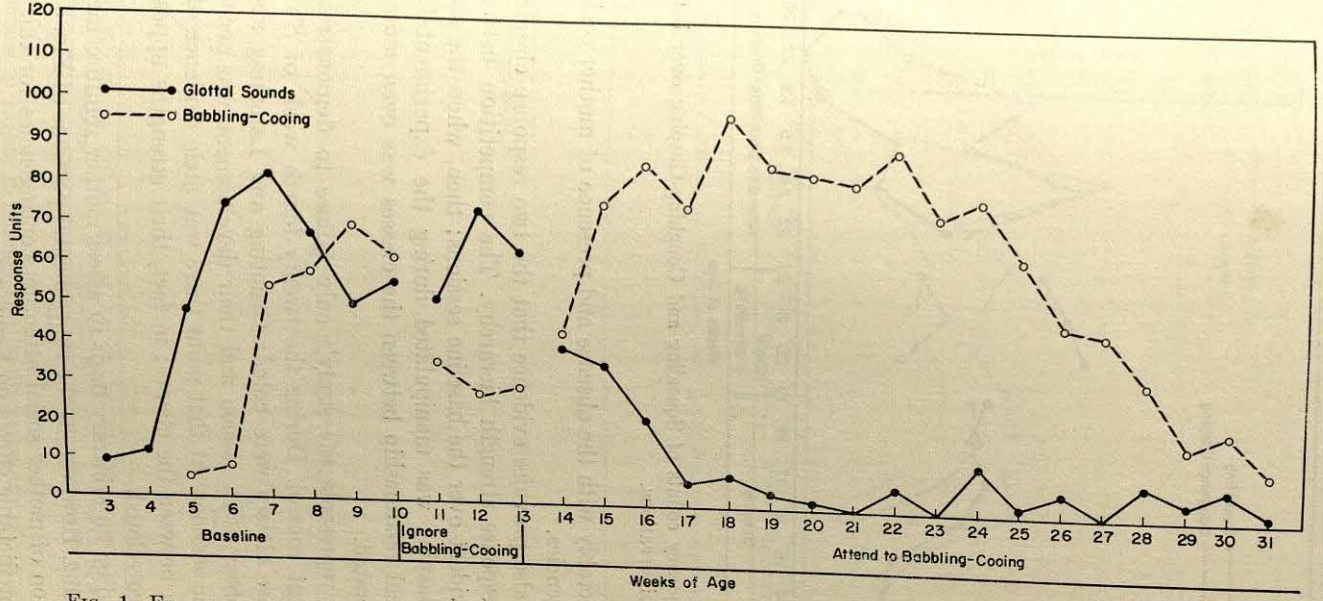


FIG. 1. Frequency counts of Glottal Sounds and Babbling-Cooing over baseline and experimental observations.

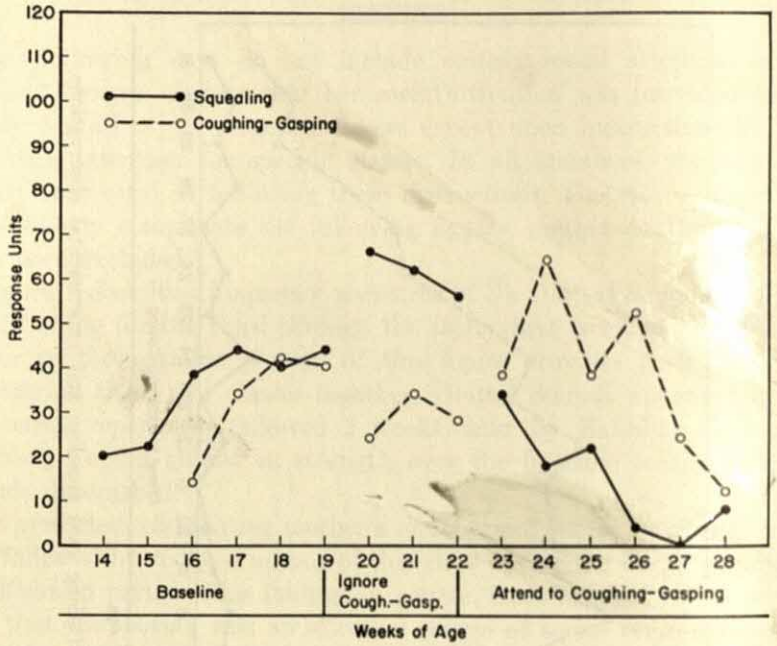


FIG. 2. Frequency counts of Squealing and Coughing-Gasping over baseline and experimental observations.

varied predictably with the absence and presence of mother social attention contingencies.

Figure 2 also provides evidence that the two response classes were in the same response strength hierarchy. The competition between these classes is evident over the baseline sessions; then, when the strength of Coughing-Gasping was manipulated during the experimental sessions, the reciprocal relationship between the classes was even more dramatically illustrated.

Figure 3 represents the study's only failure to demonstrate mother reinforcement control. During the twenty-fourth week of *S*'s life, two new response classes were noted: Grunting and Laughing appeared in the same observation session and their development was recorded over five baseline sessions. At that point there was little evidence of response competition between the classes; in fact, their strengths appeared to be positively correlated.

E decided on an arbitrary basis to assess mother reinforcement control over Laughing. However, as the experimental portions of Fig. 3 illustrate, manipulation of mother social attention contingencies for this class produced no predictable effects on Laughing.

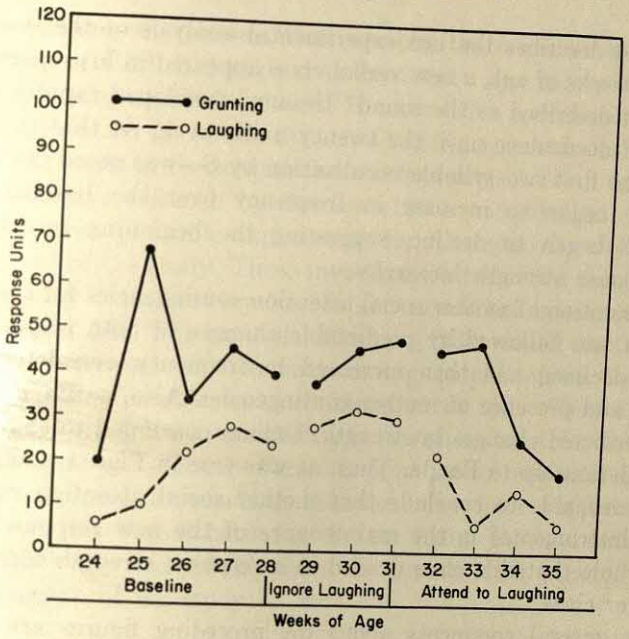


FIG. 3. Frequency counts of Grunting and Laughing over baseline and experimental observations.

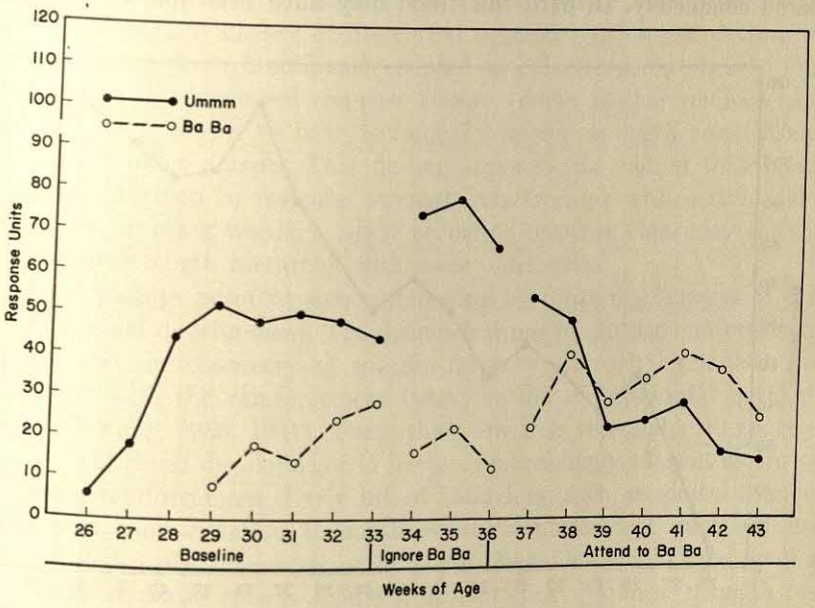


FIG. 4. Frequency counts of Ummm and Ba Ba over baseline and experimental observations.

Figure 4 describes the last experimental analysis of the study. When *S* was 26 weeks of age, a new verbal class appeared in his repertoire. This class, best described as the sound "Ummm" developed rapidly in its frequency of occurrence until the twenty-ninth week. At that time another sound—the first two-syllable vocalization by *S*—was noted; as this sound (Ba Ba) began to increase in frequency over the baseline sessions, "Ummm" began to decline, suggesting the beginning formation of a new response strength hierarchy.

Manipulation of mother social attention contingencies for the new class (Ba Ba) was followed by predictable changes in both response classes. Ba Ba declined and then increased in frequency correlated with the absence and presence of mother contingencies. Also, as Ba Ba displayed these predicted changes in strength, Ummm continued to show its reciprocal relationship to Ba Ba. Thus, as was true in Figs. 1 and 2, it would seem reasonable to conclude that mother social attention was at least partly instrumental in the maintenance of the new response class, and to conclude that this class existed in a response strength hierarchy with the older class.

Some general comments about the preceding figures are warranted. In spite of mother's continued social attention to *S*'s various response classes, all classes eventually declined in frequency and some disappeared completely. In part, this trend may have been due to response

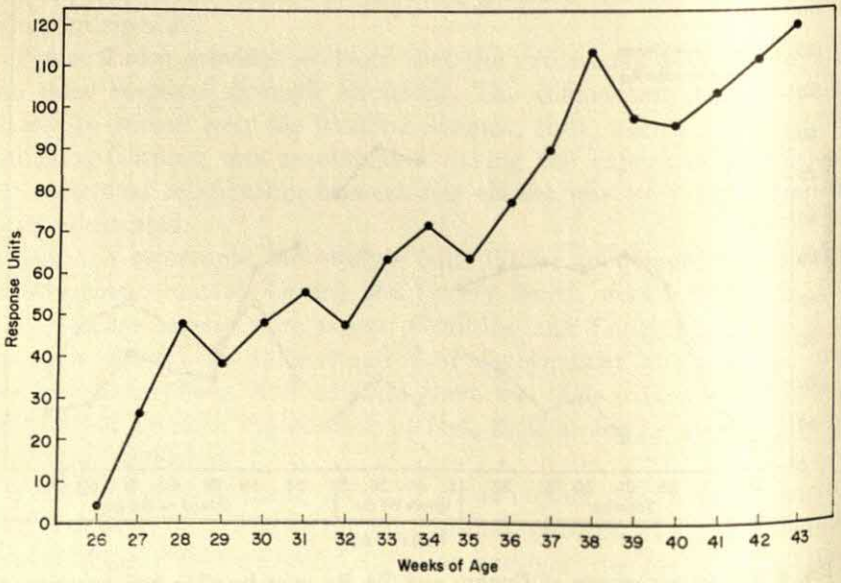


FIG. 5. Frequency counts of hand-reaching responses over 18 weekly observations.

competition among the classes. For example, the eventual decline and disappearance of Glottal Sounds may have been due to the emergence of Babbling-Cooing and later classes. However, this speculation is not a plausible explanation for the decline in all of *S*'s vocalizations from the thirty-seventh week to the termination of the study.

Early normative research by Shirley (1933) and McCarthy (1954) provides the most plausible explanation for *S*'s reduced verbal output toward the end of the study. These investigators concluded that response competition may exist between verbal and nonverbal behaviors during the first year of life. For example, Shirley (1933) found that infants' vocalizations decreased in frequency during periods when they were beginning to reach for objects, to sit alone, or to walk. Forewarned by these findings, *E* instructed the observers to also be alert for the occurrence of this kind of behavior. Figure 5 provides correlational evidence that *S*'s development of reaching behavior may have suppressed the development of his verbal behavior. As the figure shows, reaching increased dramatically and reliably from the twenty-sixth week of *S*'s life to the termination of the study.

DISCUSSION

The data reported in this study lead to two conclusions: (1) social attention provided by *S*'s mother was at least partly responsible for the control of various classes of his verbal behavior; (2) some of these response classes were functionally related to older response classes. That is, some newly developed response classes (under mother reinforcement control) were shown to bear reciprocal response strength relationships to specific older classes. This finding supports the notion that infant behaviors develop in response strength relationships with certain older behaviors; in other words, a newly occurring response class may exist in a response strength hierarchy with some older class.

These findings point to some interesting but puzzling features of this infant's social development. The demonstrations of mother reinforcement control and the discovery of specific response strength hierarchies are consistent with the reinforcement theory model of behavioral development (Gewirtz, 1956, 1961). From this view it is reasonable to conceptualize behavioral development in terms of increasingly efficient means of obtaining reinforcement. Early infant behaviors, such as glottal sounds, while they may function to obtain social reinforcement, may be less efficient than newer responses, such as babbling. One would thus predict that these response classes should bear reciprocal response strength relationships and their comparative response strengths should be under social reinforcement control.

Although these predictions were borne out during the experimental sessions, infant-mother interbehaviors during the baseline sessions did not meet these expectations. If the developing reciprocal relationships among infant response classes were due to greater effectiveness of the newer classes in obtaining mother attention, the mother's behavior should have occurred on a selective basis. That is, looking at the developmental process from the mother's standpoint, she is conceptualized from the reinforcement model as a selective dispenser of social attention; she attends selectively to successive approximations to adult behavior—weakening the more primitive elements and strengthening the more adultlike components. However, as the baseline data revealed, mother social attention was provided *continuously* for *all* infant response classes. Thus, one could not conclude that a "shaping" process was responsible for the formation of infant response strength hierarchies. It may well be that certain mother behaviors were more reinforcing for *S* than were others, but this contention could not be evaluated in the present study. No doubt, multiple stimulus classes of the mother's behavior should have been recorded as were multiple response classes of *S*'s behavior. Perhaps further studies in the area will consider this point.

REFERENCES

Bowlby, J. The nature of a child's tie to his mother. *International Journal of Psychoanalysis,* 1958, **39,** 350–373.

Bowlby, J. Separation anxiety. *International Journal of Psychoanalysis,* 1960, **41,** 1–25.

Brackbill, Y. Extinction of the smiling response in infants as a function of reinforcement schedule. *Child Development,* 1956, **27,** 205–221.

Cairns, R. B. The influence of dependency inhibition of the effectiveness of social reinforcers. *Journal of Personality,* 1961, **29,** 466–488.

Caldwell, B. The usefulness of the critical period hypothesis in the study of filiative behavior. *Merrill-Palmer Quarterly,* 1962, **8,** 229–242.

Etzel, B. C., and Gewirtz, J. L. Experimental modification of caretaker-maintained high rate operant crying in a six- and a twenty-week-old infant (Infans tyrannotearus): extinction of crying with reinforcement of eye contact and smiling. *Journal of Experimental Child Psychology,* 1967 **5,** 303–317.

Gewirtz, J. L. A learning analysis of the effects of normal stimulation, privation, and deprivation on the acquisition of social motivation and attachment. In B. M. Foss (Ed.) *Determinants of infant behavior,* London: Methuen, 1961.

Gewirtz, J. L. A program of research on the dimensions and antecedents of emotional dependence. *Child Development,* 1956, **27,** 205–221.

Hartup, W. W. Friendship status and the effectiveness of peers as reinforcing agents. *Journal of Experimental Child Psychology,* 1964, **1,** 154–162.

McCarthy, D. Language development in children, In L. Carmichael (Ed.), *Manual of child psychology* (2nd ed.) New York: Wiley, 1954.

Rheingold, H., Gewirtz, J. L., and Ross, H. W. Social conditioning of vocalization

in the infant. *Journal of Comparative and Physiological Psychology,* 1959, **52,** 68–73.

SCHAFFER, H. R., CALLENDER, W. M. Psychologic effects of hospitalization in infancy. *Pediatrics,* 1959, **24,** 528–539.

SCHAFFER, H. R., EMERSON, P. E. The development of social attachments in infancy. *Monographs of the Society for Research in Child Development.* 1964, **29,** 1–77.

SCHAFFER, H. R. Objective observations of personality development in early infancy. *British Journal of Medical Psychology,* 1958, **31,** 174–184.

SHIRLEY, M. M. The first two years, a study of twenty-five babies: Vol. II. Intellectual development. *Institute of Child Welfare Monographs.* **8,** Minneapolis: Univ. of Minnesota Press, 1933.

SPITZ, R. A. Hospitalism: An inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanalytic Study of the Child,* 1945, **1,** 53–74.

STEVENSON, H. W., KEEN, R., AND KNIGHT, R. N. Parents and strangers as reinforcing agents for children's performance. *Journal of Abnormal and Social Psychology,* 1963, **67,** 183–186.

WAHLER, R. G. Infant social attachments: a reinforcement theory interpretation and investigation. *Child Development,* 1967, **38,** 1079–1088.

WEISBERG, P. Social and non-social conditioning of infant vocalizations. *Child Development,* 1963, **34,** 377–388.

WOLFF, P. H. Observations on the early development of smiling. In B. M. Foss (Ed.), *Determinants of infane behavior,* London: Methuen, 1963.

WOLFF, P. H. Observations on newborn infants. *Psychosomatic Medicine,* 1959, **21,** 110–118.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 114–121 (1969)

Suppression of Stereotyped Screaming Behavior in a Profoundly Retarded Institutionalized Female[1]

JOHN HAMILTON AND JON STANDAHL
Gracewood State School and Hospital, Gracewood, Georgia

The study presented involved the suppression of stereotyped screaming behavior which was of such high frequency and intensity that it was very disrupting to cottage management. The method used was a shock punishment procedure which has been found effective in eliminating or reducing violent, disruptive and self-abusive behaviors in other residents. The first phase involved suppressing the screaming behavior for daily 15-minute periods during which one experimenter administered shock after every response. Next, this suppression procedure was extended to 2 half-hour periods with two different experimenters. Finally, the procedure was extended to all cottage personnel on a 24-hour basis.

In 1964 Lovaas *et al.* reported the use of painful electric shock to suppress self-injurious behaviors in two severely disturbed children. These shock contingent behaviors were suppressed within minutes and remained suppressed for months. This technique was discussed further in two subsequent publications by Lovaas and his collaborators (Lovaas, Freitag, Gold, and Kassorla, 1965; and Lovaas, Schaeffer, and Simmons, 1965). This basic procedure has been similarly used by other investigators to eliminate or drastically reduce self-injurious behaviors (Ball, Dameron, and Lovaas, 1964; Insalaco and Hamilton, 1966; Kushner, 1967; Risley, 1964; and Tate and Baroff, 1966). In no cases were undesirable side effects observed. On the contrary, a generalized improvement in behavior was reported in every case.

At this institution, the state facility for the mentally retarded, the shock punishment procedure was extended to a wide variety of serious behavior problems in severely and profoundly retarded residents. Ex-

[1] The program was aided by Hospital Improvement Project Grant MH-01722-02 from the National Institute of Health. Consistency of programming in the cottage was attributed to supervisor Lois Wilson and attendant staff Mattie Newsome, Ruby Mixson, Jeanette Tinley, Sadie Robinson, Mary Oliver, Dorothy Cook, Ellen Bridges, Evan Daniels, Willie Blackstone, Lillian Deas, Lurlia Carter, and Isabel Heath. Data collection and programming were assisted by Marilyn Goodrich. Graphing was done by Emilie Davall.

amples of these problem behaviors were: Window-breaking, paper-eating, rectal-digging, ruminating, and physically abusing oneself or other residents. These cases generally involved a definitive response which occurred persistently at least several times a day. The program involved administering a short electric shock with a three-battery powered shock stick immediately following each response. In most cases this punishment procedure was effective immediately in reducing and eventually eliminating the programmed behaviors. As reported by others, no undesirable side effects were observed.

A few cases, however, involved a high frequency, regularly occurring response which was not suppressed rapidly and which did not generalize readily to no-shock conditions. In this paper a case is presented in which screaming in the cottage was controlled by means of electric shock. This case was selected for presentation because it involved a gradually suppressed high frequency response, shock adaptation, and procedures of generalization.

METHOD

Subject

Pat is a profoundly retarded 24-year-old female who has been institutionalized for over 14 years. She has always required custodial care in dressing, feeding, bathing, and toileting. She has never made any intelligible vocalizations. Since institutionalization, her most characteristic behavior has been to emit low pitched, high intensity screams at a rate of about ten every minute during most of her waking hours. Due to its characteristic sound, this screaming behavior will be described hereafter as "growling." To attest to the distinctiveness of this growling response, a 10-minute tape recording was made in the cottage, during which Pat emitted 151 growls. Four raters independently classified each of these responses identically.

The intensity of this growling behavior was not only very disturbing to cottage personnel and residents but made working with Pat extremely difficult. Various approaches to the problem had been attempted in the past. The least systematic program involved assigning a volunteer worker to her on a regularly scheduled basis to provide attention, activity, and companionship. The most systematic program involved providing social attention and consumable rewards during non-growling periods and withholding these rewards during growling periods. All such attempts to modify the growling behavior failed. Two other types of vocalization were present which were not of a disrupting nature: A

low, soft moaning sound, described hereafter as "mooing"; and a series of soft, high pitched, sharp sounds, described hereafter as "chattering."

Apparatus

Recording apparatus. In order to record four separate events over time (growling, mooing, chattering, and shocking), a recording box was assembled with four micro-switches on top, one for each event. The box contained a 24-volt power supply, a 30-minute timer, a cutoff switch, and a four-pen Rustrak recorder with a paper feed of 1 inch per minute.

Shock apparatus. Two types of shock equipment were used. The first was a Lee-Lectronic Trainer (Lee Supply Company, Tucson, Arizona), consisting of a receiving and transmitting unit. The battery powered receiver was an 8-ounce unit worn on a waist belt with two penny-size electrodes spaced 10 inches apart touching the skin. The battery powered transmitter was a compact portable unit operated by an on-off switch. When activated, the transmitter released in the receiver a mild, medium or heavy (depending on setting) capacitor charge to the subject.

The second type of shock equipment used was a standard shock stick (Hot-Shot Products Company, Minneapolis, Minnesota); a tube 12 inches long by 1½ inches in diameter, containing two electrodes ¾ inches apart, an induction coil, an on-off switch, and three size C, 1.5-volt batteries.

The types of shock administered by these two pieces of apparatus were qualitatively different. The first was a single large pulse while the second was a rapid series of short pulses. Both delivered a painful, but noninjurious, shock.

PHASE 1

Procedure

Pre-training sessions. During half-hour sessions for an 8-day period, the frequencies of growling, mooing, and chattering vocalizations were recorded. During these base rate sessions, Pat was restrained in a sleeve jacket to a bench in the cottage. (Such restraint was a customary procedure due to her chronic habit, at that time, of rectal-digging, therefore, it did not create an unusual condition.) The shock apparatus was strapped to her waist during these sessions, but no shock was administered.

Training sessions. From sessions 9 through 26 the procedure was the same as during pretraining except that during the last 15 minutes of

each 30-minute session, a shock was administered immediately following every growl. Through session 18 the shock intensity was set at medium, but because of an apparent adaptation effect it was changed to heavy from sessions 19 through 26.

Results

Cumulative functions of the number of growls during each half-hour base rate and training session are presented in Fig. 1.

During base rate sessions, the rate of growling was high and fairly consistent, ranging from 178 to 435, with an average half-hour rate of 330. During the first two training sessions (Sessions 9 and 10), no effect on growling was evident. In Sessions 11 and 12 growling was sharply reduced during shock periods and also, to a lesser degree, during no-shock periods. In Session 13, during the first 12 minutes, growling returned to its previous level, but suppression was indicated by reduced growling near the end of the first 15 minutes which continued throughout the shock condition of the second 15 minutes. Sessions 14 through 18 reflected a return to base rate levels during both shock and no-shock periods. Apparently, the suppressive effect of medium intensity shock had lost its effect.

With the increase of intensity of shock in Session 19, an immediate suppression effect occurred in the shock condition which remained with

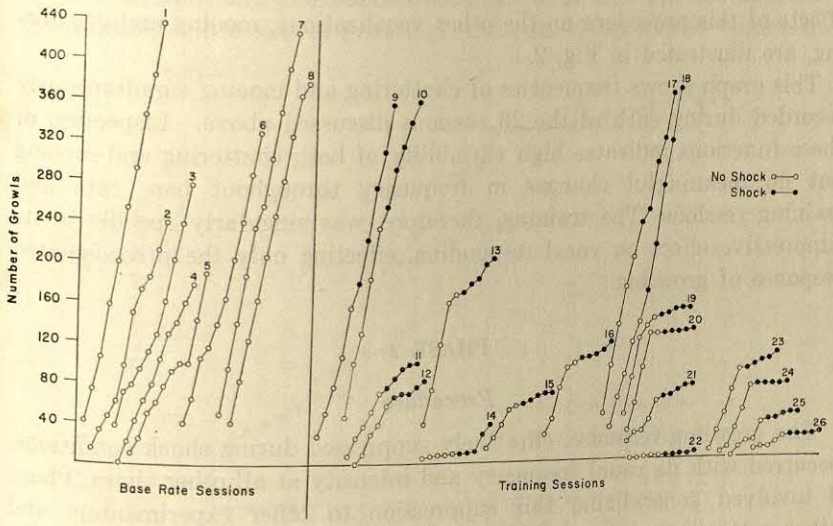


FIG. 1. Cumulative frequencies of growls in increments of 3-minute periods for each 30-minute session in Phase 1.

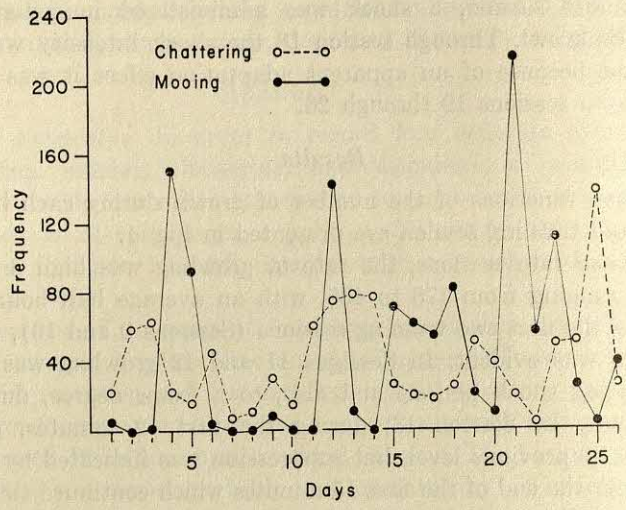


FIG. 2. Frequency of "mooing" and "chattering" vocalizations for each 30-minute session in Phase 1.

little change through Session 26, when the Phase one training was terminated. The variability in growling between sessions 19 and 26 was restricted primarily to no-shock periods.

During the shock condition growling was effectively suppressed. The effects of this procedure on the other vocalizations, mooing and chattering, are illustrated in Fig. 2.

This graph shows frequencies of chattering and mooing simultaneously recorded during each of the 26 sessions discussed above. Inspection of these functions indicates high variability of both chattering and mooing but no meaningful changes in frequency throughout base rate and training sessions. The training, therefore, was singularly specific in its suppressive effect on vocal responding, effecting only the predesignated response of growling.

PHASE 2

Procedure

The growling response, effectively suppressed during shock conditions, occurred with its usual frequency and intensity at all other times. Phase 2 involved generalizing this suppression to other experimenters and other cottage conditions.

For two half-hour periods 5 days a week, Pat was fitted with the shock receiver, but instead of being restrained to the bench she was

allowed to wander freely in the cottage or in the outdoor play yard. Two different experimenters, one was assigned to an a.m. period and the other to a p.m. period, administered shocks every time growling occurred.

Results

The results of this procedure are presented in Fig. 3.

Observation of these cumulative functions indicates that by day 18 growling was almost eliminated during these two half-hour sessions. The average growling frequencies during the first 17 morning and afternoon sessions were 25.5 and 19.2, respectively; from Sessions 18 through 32 the average rates were 2.2 and .8 respectively. Comparing morning with afternoon sessions suggests that growling may have been a stronger response in the mornings, but in both cases the response clearly was suppressed by training day 18.

PHASE 3

Procedure

The growling response, successfully suppressed during these extended training sessions, still occurred with its usual frequency at all other times. The program then was expanded to a 24-hour basis on a loosely defined intermittent punishment schedule in an attempt to facilitate a generalized suppression of growling. Azrin and Holz (1966) summarized

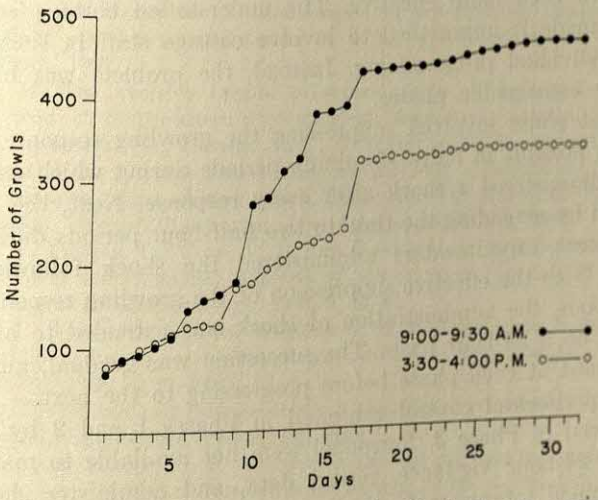


FIG. 3. Cumulative frequency of growls for the 30-minute a.m. sessions and for the 30-minute p.m. sessions during Phase 2.

the research comparing continuous and intermittent punishment schedules by stating that when the punishment contingency is discontinued, continuous punishment allows more rapid recovery of responses than does intermittent punishment. The remote control shock apparatus was replaced by the shock stick to facilitate around-the-clock programming. All cottage personnel were instructed to administer shock on the average of one out of ten growls. After several growls, the attendant would say "no," take the shock stick from the central station in the cottage, and give Pat one brief shock on the arm.

Results

Growling gradually diminished until, 3 months later, growling behavior occurred infrequently. When it did occur, it was most often at times when Pat was out of the cottage, such as in the dining room or play yard. For all practical purposes, therefore, this disturbing growling, which used to occur many thousands of times a day, was no longer a problem. After 1 year of continued infrequent growling behavior the program was discontinued.

DISCUSSION

In most cases, a specific response can be rapidly suppressed by making shock contingent on each occurrence of that response. This procedure has been found to be effective and practical. In the case presented in this paper, such a program consistently carried out on a 24-hour basis might have been more effective. The understaffed cottage conditions, however, made it impractical to involve cottage staff in such circumscribed individual programming. Instead, the problem was broken up into more manageable phases.

The first phase involved suppressing the growling response as completely as possible in daily 15-minute periods during which one experimenter administered a shock after every response. Next, the base was broadened by extending the time to two half-hour periods during which two different experimenters administered the shock following every response. With the effective suppression of the growling response during these periods, the administration of shock was extended to all cottage personnel on a 24-hour basis. The succession was gradual, and success was attained at each phase before progressing to the next.

The experimental control exhibited in Phases 1 and 2 by necessity was reduced in Phase 3. Personnel were not available to maintain an exclusive 24-hour vigilance, record data, and administer shock on a predetermined intermittent schedule. The cottage personnel had, however, been trained to execute individual and group behavior modification

programs with considerable reliability in this cottage setting. Therefore, many of the customary shortcomings of generalizing results from the experimental setting to the daily living environment were minimized.

There seems little question that the shock punishment procedure, when programmed for specific behaviors and consistently carried out, is a very effective suppressive technique. However, the advisability of introducing this technique as a control procedure to general cottage settings is questionable. Although no undesirable side effects were observed in the residents involved in shock programs in this setting, the technique itself had some undesirable consequences. First, the use of shock itself promoted social criticism from other areas. Second, the very effectiveness of the method tended to create a negative attitude in the cottage personnel toward considering alternative methods for controlling undesirable behaviors. It is therefore recommended that shock as a suppressive technique be used only (1) when the behavior to be suppressed is dangerous to the resident or others or is so disruptive that training in alternate and more appropriate modes of responding is severely inhibited, and (2) when the more positively directed forms of behavioral control have been attempted unsuccessfully.

REFERENCES

Azrin, N. H., and Holz, W. C. Punishment. In Honig, W. K. *Operant behavior: areas of research and application.* New York: Appleton, 1966, 380–447.

Ball, T. S., Dameron, L. E., and Lovaas, O. I. Control of self-destructive behaviors in mentally retarded children. Unpublished manuscript, 1964.

Insalaco, C., and Hamilton, J. Modification of self-injurious behavior through the use of punishment and reward in a free operant situation. Unpublished manuscript, 1966.

Kushner, M. The aversive control of several forms of maladaptive behavior. Paper read at Southeastern Psychological Association, Atlanta, 1967.

Lovaas, O. I., Freitag, G., Kinder, M. I., Rubenstein, D. B., Schaeffer, B., and Simmons, J. B. Experimental studies in childhood schizophrenia. Developing social behavior using electric shock. Paper read at American Psychological Association. Annual Convention, Los Angeles, California, 1964.

Lovaas, O. I., Freitag, G., Gold, V. J., and Kassorla, I. C. Experimental studies in childhood schizophrenia: analysis of self-destructive behavior. *Journal of Experimental Child Psychology,* 1965, **2**, 67–84.

Lovass, O. I., Schaeffer, B., and Simmons, J. Q. Building social behavior in autistic children by use of electric shock. *Journal of Experimental Research in Personality,* 1965, **1**, 99–109.

Risley, T. The effects and "side effects" of the use of punishment with an autistic child. Unpublished manuscript, 1964.

Tate, B. G., and Baroff, G. S. Aversive control of self-injurious behavior in a psychotic boy. *Behavior Research and Therapy,* 1966, **4**, 281–287.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 122–135 (1969)

The Heart-Rate Response of Four-Month-Old Infants to Auditory Stimuli[1]

RACHEL KEEN CLIFTON[2] AND WILLIAM J. MEYERS
Institute of Child Behavior, University of Iowa

Fourteen four-month-old infants were presented with a tone of moderate intensity, similar to one previously used with newborns, to determine if cardiac response patterns were the same in the two age groups. Stimulus repetition effects and then dishabituation to a nonfamiliar stimulus were also investigated. While the newborns had responded to this auditory stimulus with cardiac acceleration, the older infants showed deceleration. The largest response decrement occurred between the first six trials and the remaining trials. Response to the first nonfamiliar stimulus differed from response on the immediately preceding trial. It was hypothesized that the appearance of a decelerative heart-rate response in the early months of life may signify the development of the orienting response.

The typical heart-rate (HR) response of the newborn has been found to be a monophasic increase which lasts for several seconds (Bartoshuk, 1962a, 1962b; Bridger, 1961; Keen, Chase, and Graham, 1965). In contrast to the newborn's response, older infants around four to six months of age are reported to show decelerative HR responses (Kagan and Lewis, 1965; Lewis, Kagan, Campbell, and Kalafat, 1966; Meyers and Cantor, 1967). A major procedural difference between the two sets of studies is that simple auditory or tactile stimuli were used with newborns while rather complex visual stimuli were used with the older infants. The opposing direction of HR response observed in newborns as compared to older infants could be due to differences in stimulus properties or devel-

[1] This research was supported by a postdoctoral Public Health Fellowship to R. K. C. and by grant MR 11803-01 to W. J. M. from the National Institute of Mental Health. The authors thank Laura Joseph for help in data collection and analysis, and William Decker for aid in computer programming. A portion of this manuscript was presented at the 1968 American Psychological Association meetings in San Francisco.

[2] Now at University of Massachusetts.

opmental changes in mechanisms of autonomic control (Lipton, Steinschneider, and Richmond, 1966).

Lipton *et al.* (1966) presented the same airstream stimulus to the same *Ss* at three different ages and found evidence for a developmental shift in the HR response. Each *S* received at least three sessions, the first during the newborn period, and the second and third at 2½ and 5 months. At all three ages an initial acceleration was observed to the airstream stimulus. However, larger and longer accelerations were noted during the newborn period than when infants were older. Further, the infants, when older, had larger "maximal overcompensation" responses, which indicated that they showed more deceleration below the baseline following a brief initial acceleration.

The present study investigated the possibility of a developmental shift in the HR response to auditory stimuli. Four-month-old infants were presented with a 10-second tone whose characteristics matched the stimuli in a previous newborn study (Clifton, Graham, and Hatton, 1968). Out of five stimulus durations employed in the Clifton *et al.* study, a 10-second tone produced maximal HR accelerations. If this stimulus elicited a HR deceleration in the older infants, this would support the hypothesis that a developmental shift in the HR response occurs sometime within the first few months of life.

A second question investigated was whether pulse frequency of a tone would affect the magnitude of response. In addition to the constant or non-pulsed stimulus employed by Clifton *et al.* (1968), two other types of stimuli were presented by pulsing the tone on and off at different rates within the 10-second stimulus period. A pulsed stimulus can be considered more complex than a constant stimulus and it was predicted that magnitude of HR deceleration, if it appeared, would be larger for more complex tones.

A third concern was with the phenomena of habituation and dishabituation. Habituation is defined here as a decrement in responding to a stimulus which has been presented repeatedly without reinforcement. Dishabituation is the recovery of the habituated response when a new or nonfamiliar stimulus is presented. In this study, *Ss* were initially presented with the same stimulus for several trials, then a nonfamiliar stimulus was introduced. Those *Ss* receiving the constant tone on initial trials were switched to the pulsed tone, and vice versa. A decrement in the response was expected with repetition of the stimulus, while an increment was predicted to accompany the nonfamiliar stimulus. It may be that stimulus complexity affects response habituation, but the effect would be difficult to evaluate if complexity produces different initial response magnitudes, as was hypothesized earlier.

METHOD

Apparatus

The *S*s were tested in a quiet, sound-attenuated room, with recording equipment housed in an adjacent room. Two response measures were obtained, HR and skin resistance. Beckman Biopotential electrodes were taped to *S*'s chest, and the EKG was monitored on a Grass polygraph. A pulse coincident with the *r*-wave from a Fels cardiotachometer triggered a Hewlett-Packard Counter-printer which continuously printed out heart periods in milliseconds. Skin resistance was recorded from two electrodes attached to the sole and calf of the left leg, but these data were unusable due to frequent malfunction of the amplifier.

The auditory stimulus was the square wave output of an Eico audio generator set at a frequency of 300 pps. This output was fed through a speaker placed about eight feet away from *S*, and gave a reading of 70 db (re 0.0002 dynes/cm^2) on a General Radio sound level meter placed at the site of the infant's head. Stimulus onset and intertrial intervals, which varied randomly between 40 and 60 seconds, were controlled by a tape reader and Hunter timers. A Hunter recycling timer was also used when the stimulus was pulsed rather than constant. Stimulus onset and offset were recorded on the polygraph and on the printout of heart periods.

Procedure

Experimental procedures were explained to the mother, the electrodes were attached to the infant, and *S* was placed in an infant seat in a semi-reclining position. The mother and *E* watched the infant through a one-way window from the equipment room, while a second *E* sat behind and out of sight of *S*. The only contact made with *S* during the experimental session was *E*'s offering a pacifier if *S* cried. *E* remained out of sight as much as possible when offering the pacifier.

All *S*s were presented with a 10-second stimulus for 15 trials, then received a new stimulus on trials 16 through 20. Three different stimulus patterns were used. In a constant condition, *S*s heard the uninterrupted square wave during the 10-second stimulus period. In two pulsed conditions, the 10-second period was broken into either 1-second on, 1-second off, or ½-second on-off cycles. Thus, defining complexity as the number of on-off cycles or elements, the constant stimulus had only one such element, the stimulus pulsed every second had five, and the stimulus pulsed every ½-second had ten elements.

Subjects

The subjects were solicited by running an ad in the local newspaper. A total of 29 infants participated in the study. If an infant cried continuously throughout the 10-second stimulation period on three consecutive trials, the session was discontinued and the data were not analyzed. In such cases, the authors felt that any habituation or stimulus effects present would be disrupted by the infant's state. In addition, crying infants usually cried louder than the 70-db tone, thus creating a higher ambient noise level and making it difficult to hear the stimulus on those trials. Thirteen infants failed to complete the experimental session due to crying on three consecutive trials, and two more records were eliminated due to apparatus failure leaving 14 *Ss* whose data were analyzed for this experiment.

For the first 15 trials, seven *Ss* received the constant tone, (referred to as the Constant group hereafter), three received the 1-second pulsed tone, and four the ½-second pulsed tone. An analysis of variance on the pulsed groups alone indicated no significant differences attributable to the type of pulsed tone, i.e., all effects involving groups were nonsignificant. In all analyses subsequently reported, the data for the two groups were combined, and will hereafter be referred to as the Pulsed group. The nonfamiliar stimulus for the Pulsed group was the constant tone. In the Constant group, four *Ss* received the 1-second, and three *Ss* received the ½-second pulsed tone for the nonfamiliar stimulus.

Of the 13 infants excluded for crying, five had been assigned to the Pulsed group and eight to the Constant group. One *S* in the latter group was excluded before the experimental session began. Range of *Ss'* age was 16–18 weeks, with a mean age of 17.3 weeks and 17.1 weeks for Constant and Pulsed groups, respectively.

Data Reduction

For each *S* on each trial the *r-r* intervals were converted to the average *r-r* interval per second, using all fractions of intervals occurring within the second and weighting each interval proportionate to the time occupied by that interval. These average *r-r* intervals were computed for 1 second preceding stimulus onset and for 20 seconds postonset on each trial. This transformation of the data allowed a second-by-second analysis of stimulus offset effects as well as onset effects. The average *r-r* intervals were then transformed to HR in beats per minute (bpm) for each of the 21 seconds described above. For the habituation phase of the experiment, data for Trials 1–15 were blocked into five 3-trial blocks for analyses and graphing. If a record was illegible for a beat or two, the average of

HR readings on either side was substituted. In one instance a trial had more than three consecutive seconds which were illegible, and the average of the other two trials in that block was substituted.

RESULTS

Prestimulus Level

An analysis of variance of prestimulus level (HR during the second immediately preceding stimulus onset) indicated no significant differences for groups (F (1, 12) = 0.69, error MS = 1354.2) or trials (F (4, 48) = 0.50, error MS = 67.5). However, prestimulus level does vary among individual *Ss*, and it affects magnitude of response to a stimulus such that high prestimulus HR is associated with large changes while a low prestimulus rate is associated with small changes to a stimulus if the expected change is HR deceleration (Benjamin, 1963; Lacey, 1956; Wilder, 1956). The present data were adjusted for initial level effects with the following formula: $AD = D - b_{dx}\ (X - \bar{X})$, where AD = adjusted difference, D = postonset HR minus prestimulus HR, b_{dx} = —.153, and $\bar{X}$ = 152.3 bpm. In this adjustment, b_{dx} is a pooled linear regression coefficient (Snedecor, 1956, pp. 394–399. Homogeneity of regression coefficients was tested in the following ways: (1) Between groups; (2) among the first 18 trials at each postonset second (i.e., 1–20), using X = HR during the second immediately preceding stimulus onset; in addition Trials 1–18 were tested at Seconds 11–20, using X = HR on Second 10, the second immediately preceding stimulus offset; (3) the Trials × Groups interaction was tested at Seconds 1–10, using X = HR on the —1 second, and Seconds 11–20 using X = HR on Second 10. No F ratios were significant, indicating homogeneity of linear regression. Difference scores adjusted by the above formula replaced each of the second-by-second data points previously described.

Response Curves

The shape of the HR response was examined by graphing and through statistical analyses. The response curve, averaging over trial blocks for each of the 20 postonset values, was an immediate HR deceleration for both groups. In Figure 1, adjusted HR difference scores are plotted at the end of the time interval they represent for the first 20 seconds after stimulus onset. It should be noted that since the heart periods and the stimulus code were printed out at the *end* of each *r–r* interval, stimulus-onsets were recorded about ¼-second, on the average, after they occurred. Thus in all figures, all postonset seconds are indicated approximately ¼-second later than they actually occurred.

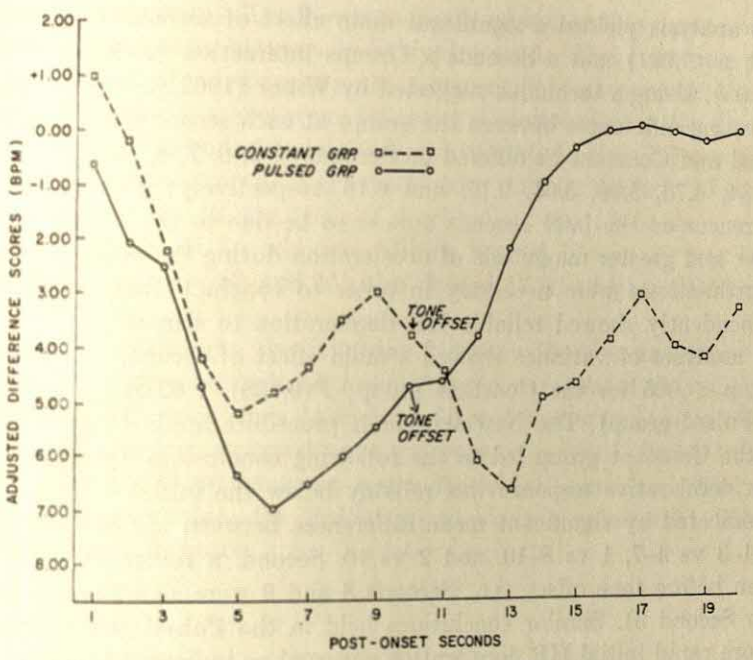


FIG. 1. Adjusted difference scores for Constant and Pulsed groups on postonset Seconds 1 through 20. Curves are means of the first 15 trials, and illustrate both stimulus onset and offset effects.

For the Pulsed group, deceleration peaked at 6 seconds postonset, then began a slow return to base level, which lasted until Second 16. The Constant group exhibited a more complicated response curve. The initial deceleration to stimulus onset was of smaller magnitude and peaked at postonset Second 5, before beginning a return toward baseline. A second decelerative wave began at stimulus offset, peaked at postonset Second 13, and remained below the response curve of the Pulsed group throughout the time period graphed.

A $2 \times 5 \times 10 \times 2$ analysis of variance was performed with one between S factor (Groups) and three within S factors (Trial blocks, Seconds, and Stimulus on-off). The on-off factor was obtained by dividing the 20 postonset seconds into Seconds 1–10 (on) and Seconds 11–20 (off). No main effects were significant, but three interactions were found: Seconds $\times$ On-off ($F(9,108) = 17.70$, $p < .005$); Groups $\times$ On-off ($F(1,12) = 9.03$, $p < .025$); and Seconds $\times$ Groups $\times$ On-off ($F(9, 108) = 3.59$, $p < .01$).

In an effort to break these interactions down into simple effects, the on-off factor was split and an onset analysis of variance (Seconds 1–10) and an offset analysis of variance (Seconds 11–20) were performed. The

onset analysis yielded a significant main effect of seconds ($F(9,108) = 10.72$, $p < .001$) and a Seconds × Groups interaction ($F(9,108) = 2.04$, $p < .05$). Using a technique suggested by Winer (1962, p. 310), contrasts for testing differences between the groups at each second were computed. Pulsed and Constant *Ss* differed on Seconds 1, 2, 6, 7, 8, and 9 ($F(1,12) = 6.34$, 4.76, 5.90, 5.63, 9.15, and 6.16, respectively; $p < .05$). These differences on the later seconds appear to be due to the Pulsed group's longer and greater magnitude of deceleration during the stimulus period.

Further tests were necessary in order to conclude that both groups independently showed reliable HR deceleration to stimulus onset. One-way analyses of variance showed a main effect of seconds ($F(9,108) = 9.80$, $p < .005$ for the Constant group; $F(9,108) = 62.04$, $p < .001$ for the Pulsed group). The Newman-Keuls procedure applied to the means for the Constant group led to the following conclusions ($p < .05$). The peak decelerative response was reliably below the initial response level as indicated by significant mean differences between the Second values for 1–3 vs 4–7, 1 vs 8–10, and 2 vs 10. Second, a return to base level began before tone offset (i.e., Seconds 8 and 9 were significantly higher than Second 5). Similar conclusions held in the Pulsed group although a more rapid initial HR deceleration occurred as indicated by differences among Seconds 1, 2, and 3. The Pulsed group showed the additional significant comparisons: Seconds 1–3 vs 4–10, and Seconds 5–7 vs 4 and 10. These findings substantiate the impression from Fig. 1 that HR in both groups decelerated to stimulus onset.

For the offset analysis, new adjusted difference scores were calculated in the manner previously described, except that X = HR on postonset Second 10, i.e., the second which immediately preceded stimulus offset. The offset analysis of variance confirms the impression from Fig. 1 that the groups differ during this phase. Significant main effects of groups ($F(1,12) = 35.09$, $p < .001$) and seconds ($F(9,108) = 8.37$, $p < .01$) were found, plus a Groups × Seconds interaction ($F(9,108) = 2.34$, $p < .05$). Contrasts for individual seconds revealed that the groups differed on all seconds ($F(1,12 \geqq 8.03$, $p < .05$). If the curve for the Constant group had not changed direction after stimulus offset, one might infer that the group difference in the offset analysis of variance was probably due to a slower return to base level for the Constant group. However, the appearance of a second decelerator wave following stimulus offset warranted further investigation to determine whether this was a reliable response. Preparatory to running comparisons between seconds, one-way analyses of variance were performed separately on each group, resulting in a significant main effect of seconds for both groups ($F(9,108) = 8.23$, $p < .005$ for the Constant group; ($F(9,108) = 19.61$, $p < .005$

for the Pulsed group). The Newman-Keuls procedure was again used for making comparisons between individual seconds. The following statistically significant comparisons are pertinent for determining the reliability of the decelerator wave following stimulus offset in the Constant group: HR on Seconds 12 and 13 was decelerated below HR on Second 11; HR on Second 12 was decelerated more than on Seconds 15 through 20; HR on Second 13, the peak of deceleration, was lower than on Seconds 14 through 20. In addition, HR on Second 17 and 20 was higher than on Seconds 11, 14, and 15; also Second 17 was higher than Second 19.

In the Pulsed group, the following Newman-Keuls comparisons were significant: HR on Seconds 13 through 20 was accelerated above Seconds 11 and 12; HR on Seconds 14 through 20 was accelerated above Second 13. All comparisons reported are significant at $p < .05$. These findings suggest that the Constant group responded to both onset and offset with HR deceleration while the Pulsed group made a substantial HR deceleration only to onset.

Stimulus Repetition Effect

In the analyses reported above, no main effect or interactions involving trials were found. Data for the five separate trial blocks are plotted in Fig. 2 for the 20 postonset seconds. Trial effects were most evident during the first few seconds after stimulus onset; response curves on later seconds showed greater variability and less difference among trial blocks. The 20 postonset seconds were broken down into 5-second segments to yield four new analyses. Trial blocks were a significant effect only in the analysis of variance on Seconds 1–5 ($F(4, 48) = 3.39$, $p < .025$). Curves in Fig. 2 indicate that HR on Trial blocks 1 and 2 decelerated below HR on the other trial blocks. Comparisons for the following orthogonal set were made: Trial blocks 1 and 2 vs the rest; 1 vs 2; 3 vs 4 and 5; and 4 vs 5. Only the first comparison was significant ($F(1, 12) = 6.43$, $p < .05$), indicating that the response began to habituate after the first six trials.

As would be expected from the overall analyses reported in the preceding section, a seconds main effect was significant for analyses of variance on Seconds 1–5, 6–10, and 11–15 ($F(4, 48) = 17.23$, 3.84, and 6.89, respectively; $p < .01$). In addition, the stimulus offset effect was again evidenced by a Seconds × Group interaction in the analysis of variance for Seconds 11–15 ($F(4, 48) = 5.49$, $p < .01$). A group main effect was the only significant source of variance in the Seconds 16–20 analysis ($F(1, 12) = 5.48$, $p < .05$). This last effect is best illustrated in Fig. 1,

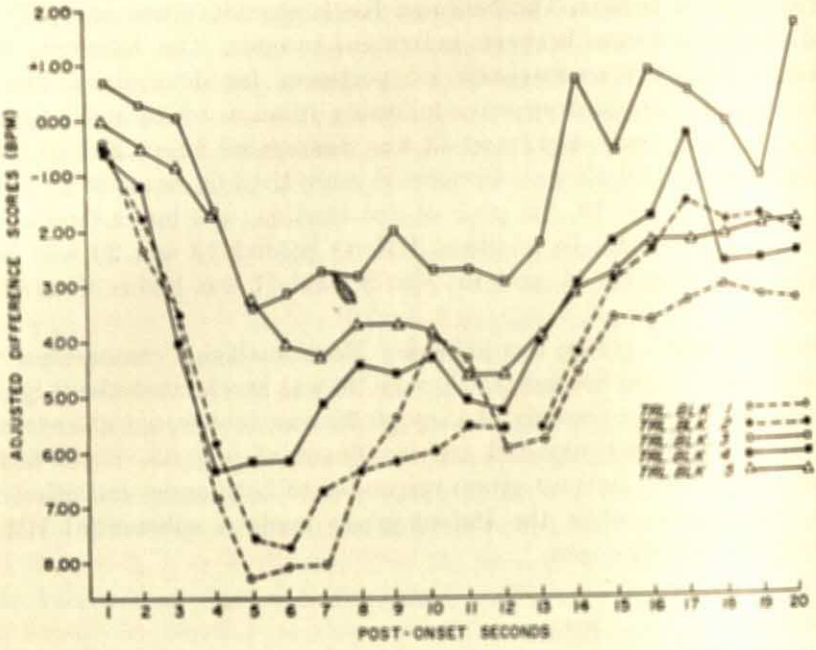


FIG. 2. Second-by-second curves for Trial blocks 1 through 5. Each curve is the mean of 3 trials and 14 *Ss*.

and appears to be due to the continued deceleration in the Constant group after the Pulsed group has returned to prestimulus level.

Dishabituation

When trials 16 through 20 were graphed separately (not shown), the effect of the nonfamiliar stimulus appeared to dissipate after one trial. To make comparisons between the habituated response and the response to the nonfamiliar stimulus, Trial 15 (the last habituation trial) was compared separately with Trials 16, 17, and 18 in six analyses of variance, three testing onset effects and three testing offset effects. Each analysis of variance contained one between *S* factor (Groups) and two within *S* factors (Trials and Seconds). The data were the same type of adjusted difference scores as were previously used except that individual trial data rather than trial-blocked data were used. The only significant outcome relevant to dishabituation, i.e., effects involving trials, was a Seconds × Trials interaction ($F(9,108) = 2.36$, $p < .05$) in the Onset analysis of Trial 15 vs 16. Individual contrasts for the various seconds showed Trials 15 and 16 to differ on Seconds 6 and 7 ($F(1, 12) = 6.27$, 7.65, respectively; $p < .05$). This appeared to be due to a monophasic deceleration on Trial 15, and a possible acceleration beginning on Second

4 for Trial 16 (Fig. 3). These apparent differences were further tested by one-way analyses of variance on Trials 15 and 16 individually. Trial 15 had a main effect of seconds ($F(9, 108) = 2.66$, $p < .05$) while Trial 16 did not. Thus, there was no significant change in HR in response to the stimulus on Trial 16, while reliable HR deceleration was observed on the previous trial.

DISCUSSION

Results of the present study indicate that four-month-old infants respond to a simple auditory stimulus with HR deceleration. This is in contrast to HR acceleration exhibited by newborns to such a stimulus. Several hypotheses might account for this directional change in the early months of life. Lipton, Steinschneider, and Richmond (1966) speculated that greater vagal tone in the older infant might produce reflex bradycardia, while the newborn's "relative deficiency" in vagal ability to produce such reflex activity might lead to HR acceleration for prolonged periods. Thus, maturation of mechanisms controlling vagal tone may explain the older infant's decelerator response.

Evidence from many sources indicates that dramatic changes are occurring in the central nervous system of the human infant between birth and six months. The increasing influence of cortical inhibition is

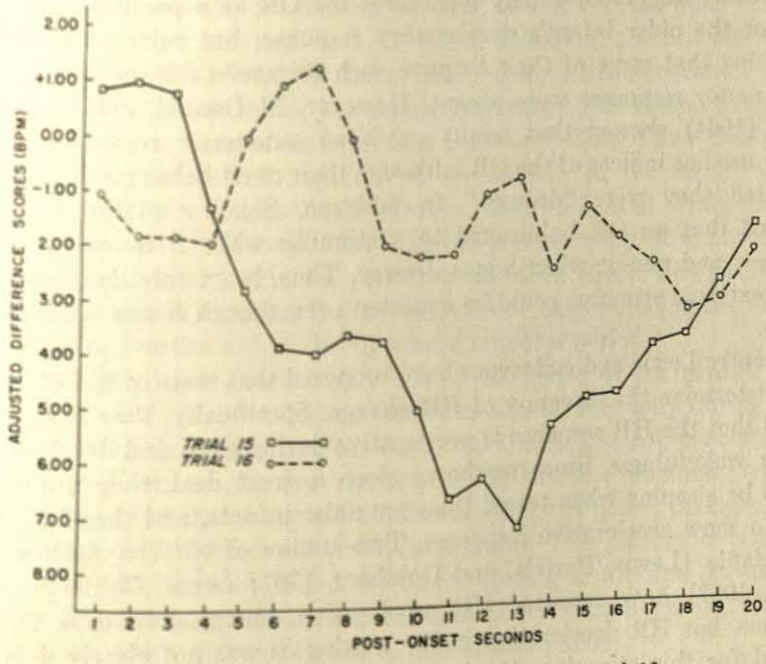


FIG. 3. Second-by-second curves for Trials 15 and 16.

involved in the disappearance of many motor reflexes, such as the Moro and the "stepping" reflex, by two to three months of age (Scheibel and Scheibel, 1964). Simultaneously, improved muscle tone, particularly in the head, neck and arms, and numerous nonreflexive behaviors, such as the social smile and visually-directed hand movements (White, Castle, and Held, 1964), become prominent. Sokolov (1960, 1963) has described a complex response system, the orienting reflex (OR), which presumably involves cortical control of the organism's response to stimulus input. One explanation for the directional shift in HR response is that the deceleration signifies development of the OR. Graham and Clifton (1966) reviewed evidence from human adult and animal research that suggested deceleration to be the HR component of the OR. Graham, Clifton, and Hatton (1968) suggested that neural mechanisms necessary for the OR may not be sufficiently developed in the newborn. These authors noted that it is too early to conclude that the newborn is incapable of a "decelerative" OR, but mounting evidence indicates that such responses are easily elicited by a variety of stimuli in the older infant, while accelerative responses are elicited by a number of different stimuli in the newborn. The question as to whether the OR as described by Sokolov is elicited in the infant can best be answered by obtaining other measures of the OR, such as the vasomotor changes in cephalic blood vessels.

Lipton *et al.* (1966) briefly mentioned the OR as a possible explanation for the older infant's deceleratory response, but rejected the idea by noting that some of their *S*s gave such responses during sleep when overt motor responses were absent. However, McDonald, Johnson, and Hord (1964) showed that adults exhibited autonomic responses commonly used as indices of the OR, although their overt behavior and EEGs indicated they were "drowsy." In addition, Sokolov (1960, p. 210) asserted that an OR habituated to a stimulus while *S* is awake, may reappear and persist when *S* gets drowsy. Thus, heart rate decelerations to an external stimulus could be expected even though *S* was apparently asleep.

Recently Lewis and colleagues have proposed that state of wakefulness may determine the direction of HR change. Specifically, they have suggested that the HR response is accelerative during sleep and decelerative during wakefulness. Since newborns sleep a great deal, they are more apt to be sleeping when tested than are older infants, and therefore will tend to show accelerative responses. Two studies of two- to eight-week-old infants (Lewis, Bartels, and Goldberg, 1967; Lewis, Goldberg, and Dodd, 1967) have reported HR acceleration during sleep to a tactile stimulus, but HR deceleration in the waking *S*s was not clearly demonstrated for this stimulus. Hord, Lubin, and Johnson (1966) presented

sleeping adult *Ss* with a 1000 cps tone, a stimulus similar to other auditory stimuli that have elicited HR deceleration in waking adults (Chase and Graham, 1967; Wilson, 1964). Sleeping *Ss* exhibited a diphasic response in which an accelerative phase lasted three or four postonset beats, and the decelerative phase lasted ten or more beats. The authors stated that the overall HR change may be described as decelerative, since the algebraic sum of the two phases is negative compared with prestimulus level. The effects of state on the infant's HR response remain unclear and should be pursued by including state as a variable in studies replicating previous stimulus conditions.

The Constant and Pulsed groups differed in two respects: The Pulsed group decelerated significantly below the Constant group to stimulus onset (on Seconds 1, 2, 6, 7, 8, and 9); and the Constant group alone showed a reliable HR deceleration to stimulus offset. It is tempting to conclude that the larger decelerations to the pulsed stimulus are due to greater complexity. Our results show differences between two levels of auditory complexity, but these differences may be due to other stimulus parameters not connected with complexity. Until *Ss'* responses discriminate a continuum using three or more levels of complexity, it would be premature to draw a conclusion relating stimulus complexity to magnitude of HR change.

The offset effect was unexpected, in that we know of no other infant studies which have reported such an effect. However, the findings of a recent study by Chase and Graham (1967) using auditory stimuli with adults parallel our result nicely. Subjects receiving an 18-second constant tone gave HR decelerations to both tone onset and offset. These decelerative responses habituated rapidly, as is common in the adult when presented with nonsignal stimuli (Davis, Buchwald, and Frankmann, 1955; Meyers and Gullickson, 1967). In the present study, it is not clear why only the constant stimulus produced an offset effect. One can suggest that the pulsed stimulus had several offsets, thus reducing the effect of the final one at the end of the 10-second stimulus period.

Stimulus repetition affected only the initial phase of the cardiac response. The Trials factor achieved statistical significance only during the first 5 seconds following stimulus onset, and the largest response decrement occurred between the first two trial blocks and the remaining trials. It should be noted that complete habituation did not occur, perhaps due to presenting only 15 trials of the same stimulus. This pattern of partial habituation, with the most rapid decrement coming in the first six trials, has been found in newborn studies using auditory stimuli (Keen, Chase and Graham, 1965; Clifton, Graham, and Hatton, 1968).

There was some evidence of a disruption in the usual deceleratory HR

response when the nonfamiliar stimulus was presented. HR during the first nonfamiliar stimulus did not change significantly during the 10 seconds following tone onset. The authors had expected a larger deceleration to the nonfamiliar stimulus than to the stimulus on the immediately preceding trial, i.e., a dishabituation effect. Although this reaction was not observed, the nonfamiliar stimulus did produce a different effect from the preceding familiar stimulus, and the effect was present for only one trial.

REFERENCES

Bartoshuk, A. K. Human neonatal cardiac acceleration to sound: Habituation and dishabituation. *Perceptual and Motor Skills*, 1962, **15**, 15–27. (a)

Bartoshuk, A. K. Response decrement with repeated elicitation of human neonatal cardiac acceleration to sound. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 68–73. (b)

Benjamin, L. Statistical treatment of the law of initial values (LIV) in autonomic research: A review and recommendation. *Psychosomatic Medicine*, 1963, **25**, 556–566.

Bridger, W. Sensory habituation and discrimination in the human neonate. *American Journal of Psychiatry*, 1961, **117**, 991–996.

Chase, W., and Graham, F. K. Heart rate response to nonsignal tones. *Psychonomic Science*, 1967, **9**, 181–182.

Clifton, R. K., Graham, F. K., and Hatton, H. M. Newborn heart rate response and response habituation as a function of stimulus duration. *Journal of Experimental Child Psychology*, 1968, **6**, 265–278.

Davis, R. C., Buchwald, A. M., and Frankmann, R. W. Autonomic and muscular responses, and their relation to simple stimuli. *Psychological Monographs*, 1955, **69**, (20, Whole No. 405).

Graham, F. K., and Clifton, R. K. Heart rate change as a component of the orienting response. *Psychological Bulletin*, 1966, **65**, 305–320.

Graham, F. K., Clifton, R. K., and Hatton, H. M. Habituation of heart rate response to repeated auditory stimulation during the first five days of life. *Child Development*, 1968, **39**, 35–52.

Hord, D. J., Lubin, A., and Johnson, L. C. The evoked heart rate response during sleep. *Psychophysiology*, 1966, **3**, 46–54.

Keen, R. E., Chase, H. H., and Graham, F. K. Twenty-four hour retention by neonates of an habituated heart rate response. *Psychonomic Science*, 1965, **2**, 265–266.

Lacey, J. I. The evaluation of autonomic responses: Toward a general solution. *Annals of the New York Academy of Science*, 1956, **67**, 123–163.

Lewis, M., Bartels, B., and Goldberg, S. State as a determinant of infants' heart rate response to stimulation. *Science*, 1967, **155**, 486–488.

Lewis, M., Goldberg, S., and Dodd, C. Cardiac responsivity to tactile stimulation in waking and sleeping infants. Paper presented at annual meeting of Society for Psychophysiological Research, San Diego, October, 1967.

Lewis, M., Kagan, J., Campbell, H., and Kalafat, J. The cardiac response as a correlate of attention in infants. *Child Development*, 1966, **37**, 63–71.

Lipton, E. L., Steinschneider, A., and Richmond, J. Autonomic function in the

neonate: VII. Maturational changes in cardiac control. *Child Development,* 1966, 37, 1–16.

McDonald, D. G., Johnson, L. C., and Hord, D. J. Habituation of the orienting response in alert and drowsy subjects. *Psychophysiology,* 1964, **1,** 163–173.

Meyers, W. J., and Cantor, G. N. Observing and cardiac responses of human infants to visual stimuli. *Journal of Experimental Child Psychology,* 1967, **5,** 16–25.

Meyers, W. J., and Gullickson, G. R. The evoked heart rate response: the influence of auditory stimulus repetition pattern reversal, and autonomic arousal level. *Psychophysiology,* 1967, **4,** 56–66.

Scheibel, M., and Scheibel, A. Some neural substrates of postnatal development. In *Review of child development research,* Vol. 1, New York, Russell Sage Foundation, 1964.

Sokolov, E. N. Neuronal models and the orienting reflex. In M. A. B. Brazier (Ed.), *The central nervous system and behavior.* New York: Josiah Macy Jr. Foundation, 1960, Pp. 187–276.

Sokolov, E. N. *Perception and the conditioned reflex.* New York: MacMillan, 1963.

Snedecor, G. W. *Statistical methods.* Ames, Iowa: Iowa State Univ. Press, 1956, Pp. 394–399.

White, B., Castle, P., and Held, R. Observations on the development of visually-directed reaching. *Child Development,* 1964, **35,** 349–364.

Wilder, J. The law of initial values in neurology and psychiatry: Facts and problems. *Journal of Nervous and Mental Disease,* 1956, **125,** 73–86.

Wilson, R. S. Autonomic changes produced by noxious and innocuous stimulation. *Journal of Comparative and Physiological Psychology,* 1964, **58,** 290–295.

Winer, J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 136–142 (1969)

Biotelemetry in Child Study

N. H. PRONKO
Wichita State University

The development of biotelemetry is traced out from its beginnings with the transistor 20 years ago and the endoradiosonde about 10 years ago. The areas studied up to this point are then reviewed, and possible applications of the technique are considered in relation to child study.

When Spelt (1948) conducted his experiment on conditioning of the human fetus *in utero*, he had to resort to recording by the only means available to him at the time, namely, a mechanical method utilizing rubber diaphragms attached to the mother's abdomen. A kick of the fetus was transmitted via the mother's abdomen to the rubber diaphragm, to the enclosed column of air within the attached tube, and finally to a rubber diaphragm attached to the other end of the tube. A writing pen resting on the latter diaphragm and adjusted to show changes in pressure in the enclosed air column recorded such changes on a rotating piece of glazed paper. What a far cry from today's possibilities!

At the present time, replication of Spelt's experiment could be achieved by means of a simple capsule resting on the mother's abdomen. It would contain a microminiature radio transmitter powered by a pen light battery. A cheap transistor radio at some distance from the mother would receive the information and a tape recorder could store it for later analysis. All this would be carried out without a lot of cluttering hardware and, especially, without cumbersome wires to cause what someone has called psychic trauma. The wireless age has finally arrived. Biotelemetry is a ready and willing handmaid to psychology.

IT ALL BEGAN WITH THE TRANSISTOR

Elsewhere, the writer (Pronko, 1968) has traced out, in greater detail than is here intended, the developments that led to the present state of the art of biotelemetry. Suffice it to say that the technological evolution leading up to it did not begin until the transistor was invented in 1948. It was this tiny, simple device that made possible the elimination of

bulky radio tubes and paved the way for the degree of microminiaturization that biotelemetry required.[1]

The next step forward was taken with Stuart Mackay's (1957) adaptation of the transistor into what would, in today's "mini" jargon, be christened a minitransmitting station but which Mackay named an endoradiosonde. The name was suggested by the radiosonde, the sensing radio-transmitting unit used in the balloons sent up in meteorological study. The prefix endo- seemed appropriate since the miniature capsule was designed for implanting inside the organism for the purpose of monitoring physiological functions.

SUBJECTS STUDIED UP TO THE PRESENT TIME

With apparatus commonly the size of an aspirin tablet, the transmitting sensor sends continuous information on any physiological measure desired. Power is supplied by a self-contained battery or power from a source outside the organism and, lately, there is talk about using a physiological source of power. Pressure and temperature have been simultaneously studied from the beginning. Gastric activity is another possibility. The capsule monitoring this function has been fed to such animals as dolphins and tortoises. The capsule involved can be left free to travel through the gastrointestinal system or it can be tethered to the animal's tooth by a fine nylon thread and held in place where desired. In this way, Mackay stumbled on to an interesting fact about a lizard that contained an endoradiosonde in its gut. The apparatus was transmitting normally until some humans entered the laboratory, whereupon "broadcasting" stopped instantly. Here was a first hint of the usefulness of the endoradiosonde in psychological inquiry.

Soon vast fields were open to much more refined study than previously possible and *at a distance*, and, more important still, without the complications introduced by the presence of the human observer. EKG, EMG, EEG, pH, blood pressure, gastrointestinal pressure, and movement are among the functions that have been brought under continuous scrutiny, recording, and storage for later analysis by computer or otherwise. Migration of the grizzly bear and its habits and hibernation have been studied by the Craigheads (1966). Movements of chimpanzee and gorilla troops can be studied by ethologists as a welcome substitute to the heroic but frustrating attempts at tracking them visually.

[1] Partial list of manufacturers of microsystems: Biocom Incorporated, 5883 Blackwelder Street, Culver City, California; Medical Systems Corporation, 43 Plymouth Road, Great Neck, New York 11023; *Science's* Guides to Scientific Instruments; Whittaker Corporation, 12838 Saticoy Street, North Hollywood, California 91605.

A study by one of Mackay's students, Carter Collins (1967) sounds like science fiction but is an actuality. It was made possible by a pinhead-sized transmitter implanted in a rabbit's eye designed to measure intraocular pressure. The purpose of the study was to learn something about the changing pressure of the vitreous humor in relation to glaucoma. As he proceeded in his investigation, Collins learned that the intraocular pressure would increase by as much as 10 mm Hg to gentle auditory, visual, tactual, olfactory, or temperature stimulation (cf. Fig. 1). Presentation of a female bunny to a male rabbit produced a similar rise. Since then, the world is waiting to learn if glaucoma results from looking too much at bunnies.

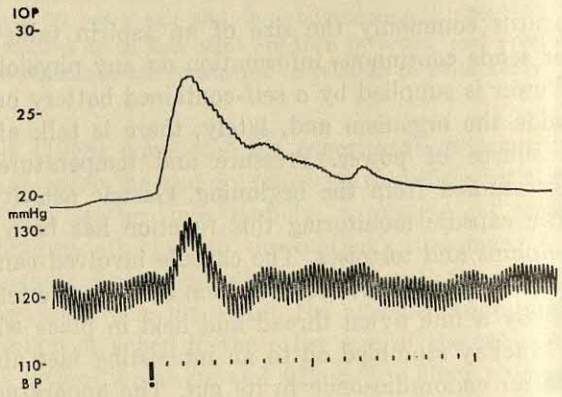


FIG. 1. (Top) Intraocular pressure (IOP) response evoked in a rabbit by a tactile stimulus (touching whisker) at arrow. (Bottom) Blood pressure (BP) measured simultaneously in the abdominal aorta. Pulsatile variations in both curves were reduced by 0.5-cycle/sec low-pass filters. Presentation of female rabbit evoked similar response. Collins, C. C. Evoked pressure responses in the rabbit eye. *Science,* 1967, **155,** 106–108.

IMPLICATIONS OF BIOTELEMETRY FOR CHILD STUDY

Fetal heart beat. Beginning close to the start of life, Hess and Litvenko (1964) have successfully monitored the fetal heart beat as well as the mother's by a wireless apparatus placed on the mother's abdomen. The technique permits study of the relationship of heart action between mother and child under various conditions of gestation, labor, and drug action. Of prime importance to the psychologist would be the effect of maternal anxiety, depression, and upset upon the fetal heart. Is some of the unique autonomic patterning of the infant, as discovered at the Fels Research Institute, to be traced to this stage? Might research via biotelemetry throw light on this question and the relation of such findings

to the infant's later temperament or other personality development? A technique is at hand for asking such questions whatever the answer might be.

Applications to child training. The ingenuity of the application of radio to behavioral research and training is largely in the offing, but an illustration of the possibilities is available in a transistorized signal package for toilet training of infants as reported by Van Wagenen and Murdock (1966). The device (as illustrated in Fig. 2) is the size of a

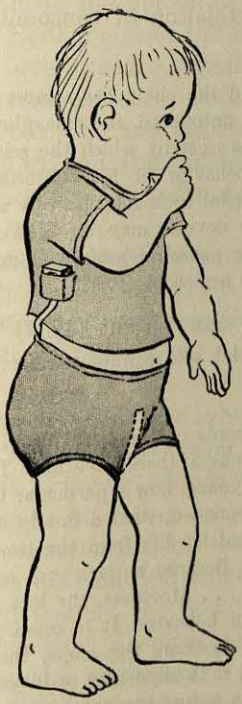


FIG. 2. Positioning of the signaling device showing the grid in exaggerated size (Van Wagenen and Murdock, 1966).

pack of cigarettes and is, in fact, contained in a polyethelene cigarette case; it weighs 4.5 ounces and places almost no limitations on the child's movements. When the child produces an expulsion, a small speaker produces a clear tone signal that is audible throughout an average-sized home. The tone signal sounds when urine or moisture from fecal matter closes the speaker circuit by making contact with a grid sewed into the child's training pants.

The apparatus has been tried on both a male and a female subject of about 16 months of age and was found to function well. The authors employed a discrim-

ination hierarchy based on proximity to the toilet facility. That is, the infant was initially reinforced at the occurrence of any expulsion, but on subsequent urinations or defecations, movement in the direction of the toilet was reinforced with confections, and the most distal areas of the house successively removed from the possibility of reinforcement. Aspects of the response as movement toward the toilet and restraining of the expulsion were the focus of reinforcement (p. 314).

Ecological research and applications. Schwitzgebel (1964) and his associates have suggested some possibilities of what they have called behavioral electronics[2] (p. 233). The following terse statement reflects their orientation and appreciation of concomitant threats of the new technique:

"Between the laboratory and the clinic lies almost the entire natural environment of humanity, practically untouched and unexplored by direct scientific procedures." Yet electronic devices exist by which the scientist could not only record but modify the day-to-day behavior of his subjects or patients. The obvious dangers of abuse of these powerful tools may be why scientists have not made use of them. But these electronic devices may be used for constructive as well as destructive purposes, and their potential usefulness in understanding and solving human problems should not be neglected (p. 233).

Some years ago, Barker and Wright (1954) expressed a need and a wish for the technique under examination here. Here is their prophetic statement:

Although we have daily records of the behavior of volcanoes, of the tides, of sun spots, and of rats and monkeys, there have been few scientific records of how a human mother cared for her young, how a particular teacher behaved in the classroom and how the children responded, what a family actually did and said during mealtime, or how any boy lived his life from the time he awoke in the morning until he went to sleep at night. Because we lack such records we can only speculate on many important questions. . . . Moreover, the lack of field data limits the discovery of some of the laws of behavior. It is often impossible to create in the laboratory the frequency, the duration, the scope, the complexity, and the magnitude of some conditions that it is important to investigate. . . . This should not be discouraging. Experiments in nature are occurring every day. We need only the techniques and facilities to take advantage of them (pp. 2-3).

It would, indeed, require a magic crystal ball to predict the possibilities and applications of behavioral electronics to child study. No one stood by

[2] Perhaps a rose by any other name smells as sweet, but it seems that behavioral scientists will have to settle on a name for the latest available electronic technique. Just as a conditioned response points to something different than an unconditioned response, so will a technique in the service of psychology have to be differentiated from biotelemetry. Since Mackay works in medicine, his preferred term is biomedical telemetry. Because "behavioral electronics" is cumbersome, would "psychotelemetry" be more advantageous?

the first Skinner box and foretold all the research and applications that this "brain child" would foment. Surely, "psychotelemetry" will develop situationally from one to another specific instance. But some uses can be anticipated by a realization of the advantages to be gained by a system that permits wholesale controls where they have been nonexistent. For example, radio transmission and tape (or other) recording of the data eliminate the presence of the human observer and his influence on the behavior under investigation. Gone are the heavy hardware and wires connecting the organism and apparatus and their unknown effect on the behavior studied. Neither should the freedom of movement permitted the organism be overlooked. Contrast the organism, child or adult, with all the harness attached, undergoing electromyographic study now and before. Under the new dispensation, a tiny capsule attached to the skin will do the job. Only controlled studies, with the old and new techniques, disclose the full advantages of the latest blessing conferred by the wireless age.

Wherever mechanical methods have been used, there, can behavioral electronics do a better job. The writer can see its applicability in all the latest and intensely exciting research in earliest infancy. Head turnings, sucklings, other movements, and perceptual responses are candidates for such study.

Ecological studies of the preschool and school child are also possibilities. Each child can have his own transmission frequency. His movements can be traced as effectively as the Craigheads were able to track individual grizzlies. A capsule can be fitted even into the heel of his shoe (Herron and Ramsden, 1967), transmitting his whereabouts as distinctively as his fingerprint identifies his biological uniqueness. Furthermore, since audio transmission is also feasible, the infant's and the child's crying, laughing, shouting, soliloquizing, dialogue, discussion, participation in play and games are all capable of continuous monitoring. What breathtaking research vistas open up with this most recent psychology's ally!

Possible applications are harder to conjure, but some of the more easily apparent ones bob up. Certainly, operant conditioning can be expected to show a population explosion of applications simply because one will not be limited to the laboratory. The home, the neighborhood, the playground, and the school room will be available for modification of behavior *in situ*, in its natural context. And, finally, the opportunities for basic research that should open up are staggering. As hinted above, we can know what children say and do to each other and with their teachers and mothers and fathers. With controlled and ethical use, the newer techniques should yield knowledge of the order achieved by the

young ethologists when they moved out of the zoos into the African jungles and mountains to study gorillas, chimpanzees, monkeys, and baboons in nature.

REFERENCES

Barker, R. G., and Wright, H. F. *Midwest and its children: The psychological ecology of an American town.* Evanston, Illinois: Row, Peterson, 1954.

Caceres, C. A. (Ed.) *Biomedical telemetry.* New York: Academic Press, 1965.

Collins, C. C. Miniature passive pressure transensor for implanting in the eye. IEEE *Trans. BioMed Eng.*, 1967, **BME-14**, 74–83.

Craighead, F., Jr., and Craighead, J. Trailing Yellowstone's grizzlies by radio. *National Geographic*, 1966, **130**, 252–267.

Gruenberg, E. L. (Ed.) *Handbook of telemetry and remote control.* New York: McGraw-Hill, 1967.

Herron, R. E., and Ramsden, R. W. A telepedometer for the remote measure of human locomotor activity. *Psychophysiology*, 1967, **4**, 112–115.

Hess, O. W., and Litvenko, W. Radio transmission of fetal electrocardiograph and intrauterine pressure. *Connecticut Medicine*, 1964, **28**, 427.

Mackay, R. S. Endoradiosonde. *Nature*, 1957, **179**, 1239–1240.

Mackay, R. S. *Biomedical telemetry.* New York: Wiley, 1968.

Pronko, N. H. Biotelemetry: Psychology's newest ally. *The Psychological Record*, 1968, **18**, 93–100.

Schwitzgebel, R., Schwitzgebel, R., Pahnke, W. N., and Hurd, W. S. A program of research in behavioral electronics. *Behavioral Science*, 1964, **9**, 233–238.

Slater, Lloyd E. (Ed.) *BioTelemetry: The use of telemetry in animal behavior and physiology in relation to ecological problems.* Proceedings of the Interdisciplinary Conference, New York, March 1962. New York: Macmillan, 1963.

Spelt, D. K. The conditioning of the human fetus *in utero*. *Journal of Experimental Psychology*, 1948, **38**, 338–346.

Van Wagenen, R. K., and Murdock, E. E. A transistorized signal-package for toilet training of infants. *Journal of Experimental Child Psychology*, 1966, **3**, 312–314.

Van Wagenen, R. K., Meyerson, L., Kerr, N. J., and Mahoney, K. Electronics, learning principles in toilet training. Mimeo.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 143-152 (1969)

The Design of Mobile Laboratories for Behavioral Research with Children[1]

KEITH G. SCOTT
University of Illinois

The design, specification, and cost of trailer laboratories for research with children are discussed. A model set of plans and specifications are presented and information about heating, towing, and insurance is provided.

During the past years a number of mobile trailers have been designed and purchased by child researchers at the University of Illinois. We have gradually evolved detailed specifications and designs that minimize costs and provide vehicles that meet both the peculiar problems of the laboratory and of being parked in school yards or apartment complexes. The ideas presented below supplement information contained in earlier publications (Bijou, 1958; Bergman, 1964).

The need for trailers is evident to most child researchers in the crowded conditions produced by an expanding area. They are frequently a necessity. Public schools, where research and testing space simply does not exist, can cooperate if a trailer is available. When a population of infants is needed the vehicles can be parked in apartment complexes. Large units can be employed as experimental classrooms for psycho-educational problems. In an institution for the retarded it is possible to establish laboratory facilities without delay when new buildings would take years to construct. At the Lincoln State School the plans for new buildings include ramps and utility hookups so that trailers may be parked where they are needed on the campus that spreads over several miles. The cost of providing laboratory space in these buildings, with its partial utilization, would be prohibitive.

Another advantage of trailers is that they provide such specialized research situations as one-way vision glass, sound control, and provision for specialized circuitry. This is not always possible to arrange even

[1]The preparation of this paper was supported by Grant MH 07346 from the National Institutes of Health. The author wishes to acknowledge the advice of his colleagues in the child area at the University of Illinois in preparation of this paper. The support of the University of Illinois Research Board in providing funds to purchase several trailers is gratefully acknowledged.

when other space is available. As a project develops, alterations to interior walls, which are of wooden construction, can be made without the great expense and trouble involved in almost any school or institution building.

The cost of a trailer per unit area varies with the size, fixtures, finish, and method of purchase. Our costs have varied between $11 per square foot for the 35×10 ft unit (Fig. 1) and $16 per square foot for the

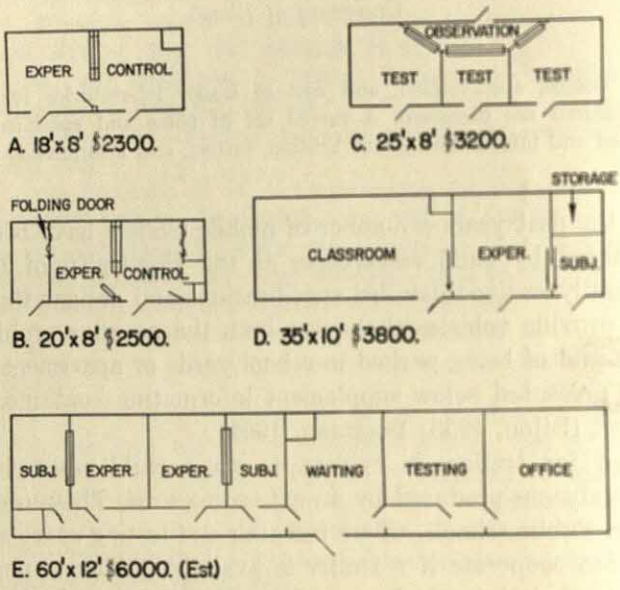


FIG. 1. Floor plans for several trailers and their costs at time of delivery. A current bid for a unit as shown in B is $2,800. Costs vary with geographic area and quotations should be sought before requests are made for funds.

20×8 ft unit. Using the typical purchasing procedures, university bids should be sought directly from companies[2] that specialize in the construction of field offices for construction and engineering projects. These firms build many one-off units with custom interiors. The particular specifications you request would disrupt the rigid production facilities of the mobile home builder who would typically charge more. You also avoid the large margin of profit added by a retailer.

An alternative to use of a trailer is the purchase of a mobile home or van with its own motive power. Experience with a 26-foot Dodge Mobile Home containing a special interior showed the vehicle to be particularly easy to drive. It was equipped with its own motor driven power plant making it independent of other supplies. Such power plants are noisy and

[2] Suppliers in the Midwest include: Mobile-Office Inc., 4845 W. 111th St., Chicago, Illinois, and Portable Structures, Box 38, Apple Road, Osceola, Indiana.

tend to provide poor voltage regulation, making them much less advantageous than might be expected. In interior space for research it was approximately equal to a 20×8 ft trailer at a cost in 1964 of $10,800. When very frequent moving is the expectation, such vehicles are highly convenient. Their high cost and relatively restricted interior space will dictate the use of trailers by most researchers.

FLOOR PLANS

Different floor plans will be required for different research purposes. Figure 2 shows the layout of a unit for laboratory learning studies. The

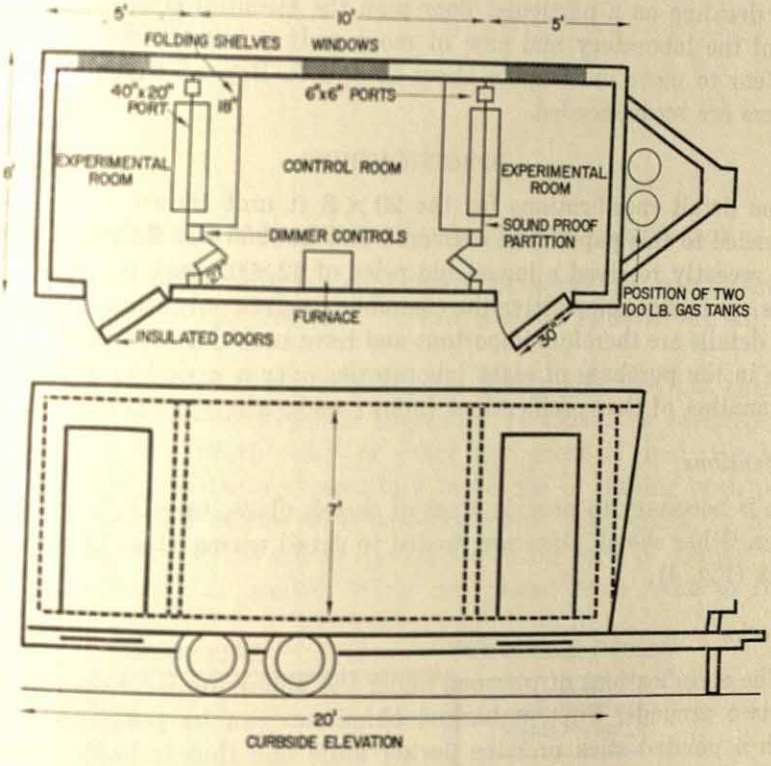


FIG. 2. Plans for the unit specified in the text.

experimental rooms are large enough to hold a projection tunnel for discrimination learning research or a marble-dropping type of equipment used on social motivation studies. By replacing the window sash between the control and experimental rooms with a blank panel, one end room provides minimum but sufficient space for psychometric testing, or, two experiments can be conducted simultaneously. A minimum cost unit is shown in Fig. 1A. It consists of a simple experimental and control room. A unit for infant research is shown in Fig. 1B. As well as space

specially built for relay racks, this has a folding door behind which a mother can wait while her infant is being tested. The unit shown in Fig. 1C was designed for supervision of students during psychometric practicum. It allows simultaneous observation of three testers by supervisors under excellent conditions. A larger unit for psycho-educational research is shown in Fig. 1D. This features a classroom area for research with teaching machines plus a laboratory area for more basic studies. Finally, Fig. 1E shows a proposed large unit for work in an institution where the laboratory would be moved only a few times a year and where a full-time staff would be employed.

In deciding on a particular floor plan the essential considerations are use of the laboratory and ease of moving. It is not too difficult for an amateur to move units up to about 25 × 8 ft. Beyond that professional movers are recommended.

SPECIFICATIONS

The detail specifications for the 20 × 8 ft unit shown in Fig. 2 are appended to this paper. The delivered cost in 1965 was $2,500. A similar unit recently received a lowest bid price of $2,800. Since prices on these units are highly competitive the customer receives only what he specifies. The details are therefore important and have been developed with experience in the purchase of eight laboratories over a period of years. Some explanation of the specifications follows under the appropriate headings.

Dimensions

It is necessary to provide a set of sketch plans, as in Fig. 2, with the order. Other sketch plans are needed to detail wiring (Fig. 3) and partitions (Fig. 4).

Exterior

The specifications of plywood under the aluminum siding is important on two grounds. First, unbacked aluminum can be punctured readily with a pointed stick or large pocket knife and thus is liable to vandal damage. Second, the backing makes the vehicle much quieter, from buffeting during high winds, and from playground noises.

Interior

The interior specifications aim to provide durable materials at reasonable cost. Birch is the least expensive hardwood paneling. Acoustical paneling is typically not durable enough to withstand the day-to-day wear and tear in such confined areas. On a wet day the floors become extremely dirty with the traffic of children coming and going. If carpet

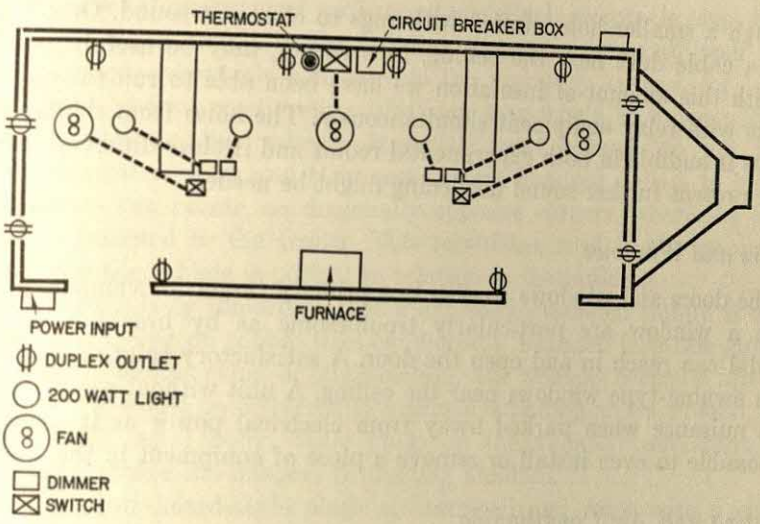


FIG. 3. Elevation of the partition for the unit specified in the text.

tempts you to reduce noise, use outdoor-indoor carpet that can be simply cut and laid over the linoleum. It can also be removed at will for cleaning.

Sound between the rooms is markedly attentuated but not entirely eliminated with the construction suggested. It should be adequate for most purposes for which relay or other equipment is used. We have found the most satisfactory procedure to be one of leaving open ports to which removable sashes with one-way vision, clear glass, or blank panels can be fitted. This allows special projection ports or display panels to be provided as needed. Wires are passed from room to room

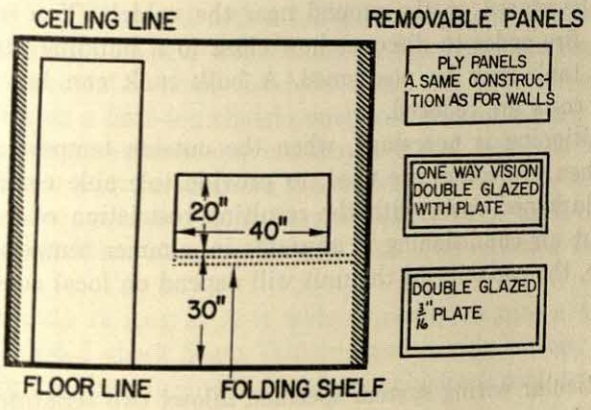


FIG. 4. Wiring diagram for the unit specified in the text.

through a smaller hole, stuffed with rags to close out sound. On a large unit a cable duct near the ceiling, in a corner, may be useful.

With this amount of insulation we have been able to run two experiments with relay equipment simultaneously. The noise from shutter and relays is audible in both experimental rooms and if clear differential cues were present further sound deadening might be needed.

Doors and Windows

The doors and windows constitute a primary target for vandals. Doors with a window are particularly troublesome as by breaking them a vandal can reach in and open the door. A satisfactory solution has been high awning-type windows near the ceiling. A unit without any windows is a nuisance when parked away from electrical power as it is nearly impossible to even install or remove a piece of equipment in the dark.

Heating and Air Conditioning

A furnace adequate for the climatic conditions is necessary and large enough to have quick recovery so that if the vehicle is moved on a cold day a wait of hours does not occur while the laboratory and equipment warms up. Heat by propane gas is recommended as adequate power outlets for electric heat cannot be found in most older school buildings. Also you pay for the propane and thus do not impose on your hosts who may not appreciate high electrical bills. Most gas supply companies will provide 100-pound bottles. If you are in a remote area purchase 40-pound bottles, which are easily carried, and take them to be filled in a car (the 100-lb bottles are too big for this). Large trailers may need four bottles unless, they are replaced very frequently in cold weather. When a laboratory is parked for an extended period a bulk tank (100–500 gallons) may be placed on the ground near the vehicle. You will need to check local fire codes to discover how close to a building (typically 15 ft) such a tank may be positioned. A bulk tank can last most of a winter and costs are reduced.

Air conditioning is necessary, when the outside temperature exceeds 70° and when projectors are used, to provide tolerable comfort. If you require a darkened room with the resulting restriction of ventilation a unit without air conditioning is unusable in summer temperatures. Like the furnace, the capacity of the unit will depend on local conditions.

Wiring

The particular wiring system specified allows two separate drop cords from the laboratory to be run up to a building. Each cord is connected

to a standard three-pronged wall receptacle. Each receptacle must be on a separate fused circuit in the building when the air conditioner is in use. No single receptacle can provide the necessary current for lights, a projector, and air conditioning. The ideal way would be to use a 220-V supply and split the voltage. However, most schools do not have such electrical services and they cost several hundred dollars to install. We provide two points, on diagonally opposite corners, where the cords can be connected to the trailer. This minimizes cord length no matter which way the vehicle is parked in relation to the building.

It is particularly important that a carefully grounded supply is used. A trailer standing on rubber tires can become insulated from the ground. If, through wiring error or malfunction, the vehicle becomes alive, a severe shock would be delivered to a person, standing on the ground, who touched the metal body. Male plugs on the vehicle, well above the ground, minimize the dangers to playing children.

These motor board-style plugs are recessed and come with a spring loaded cover that can be secured with a small lock. The particular split of the circuits given in the specifications is one we have found useful. In winter, when the air conditioner is not in use and only one projector is operating, it is possible to jump the exterior plugs and use only a single power lead. In some applications a separate circuit for the air conditioner could be installed and a lead for it run out only when it is in use. Another useful feature preferred by some is to install take-up reels, for the cables, in the trailer body.

When seeking bids for the trailer a plan (See Fig. 3) showing the position of the electrical fittings is needed.

Towing

A mobile laboratory up to 8×25 ft will weigh in the region of 4000 lb and can be moved with a full-sized sedan car with helper springs or an equalizing hitch and connections for the lights and electric brakes. This does represent a heavy load for a car and a better solution is to use a vehicle on a half-ton chassis, such as a pickup, a van, or carry-all. The ground crews at most universities have such vehicles. The hitch on the tow vehicle and the hitch on the trailer need to be of similar height for smooth towing. If several trailers are likely to be built over a period some standardization is called for and should be reflected in the specifications.

Units over 25 ft long or 8 ft wide typically require a truck to tow them. You should check State Vehicle Laws in this matter. Illinois law, for example, requires dual rear wheels on vehicles towing trailers greater than 8 ft wide.

Insurance

To insure the vehicle you must check university regulations and if a federal grant is involved, seek special permission to use funds from the agency for the premium. Collision damage ($100 deductible) and comprehensive physical damage on an actual cash value basis costs $62 per year in central Illinois for a 35 × 10-unit.

MODEL SPECIFICATIONS FOR USE IN SOLICITING BIDS

I. Dimensions (See Fig. 2).
   A. Length of body 20 ft.
   B. Width of body 8 ft.
   C. Interior height 7 ft from floor to ceiling. The floor is to be flat without intruding wheel arches.

II. Exterior
   A. Prefinished aluminum siding .024–.026 inches with backed-on prepainted white enamel finish, over ¼-inch plywood panel.
   B. Siding attached over ¼-inch plywood to 2 × 2-inch joists on 16-inch centers crossbraced at 2-ft intervals, 2 × 4-inch framing at corners, doors, and windows.
   C. Roof of 1 piece 28-gauge galvinized steel.
   D. Rafters 2 × 6 inches on 16-inch centers.
   E. Three all-weather vents with electric fan, one each in the two experimental rooms and the center control room.
   F. Insulation of 1½-inch Fiberglas with vapor barrier in all four exterior walls and 4-inch Fiberglas with vapor barrier in ceiling and floor.

III. Interior
   A. All interior walls covered with prefinished ¼-inch birch paneling.
   B. Ceiling of white acoustical tile over backing board.
   C. Floor of ⅛-inch vinyl asbestos tile with a 150-lb. per square inch rating; color of tile—neutral gray.
   D. Baseboard of rubber molding 4 inches high.
   E. Double-wall interior partitions between center control room and end experimental rooms (2 × 4 studs) to be insulated with 1½-inch Fiberglas insulation and 2-inch polyurethane foam for sound control.
   F. Openings to be left in each partition as specified in Fig. 4 attached, including a 20 × 40-inch port and a 6 × 6-inch opening in each partition. The smaller opening will be covered by a panel on each (See Fig. 4).

IV. Doors

A. Two exterior aluminum doors without windows 30 × 72 inches. The locks on doors 1 and 2 to be keyed the same. Each exterior door to have weather stripping, a pull-out step, and additional steps specified in Section IX, G.

B. The two interior doors are to be 27 inches wide and sound-insulated with bronze weatherstrip around. Threshold to be installed below door.

V. Windows

A. Three windows mounted near ceiling, one in each room on the side of the trailer opposite the exterior doors. Each window to be approximately 18 × 12 inches, awning type to open with crank mechanism. Windows to have drip caps.

VI. Heating and Air Conditioning

A. A forced air-furnace of at least 30,000 BTU to be located in control room as in Fig. 2, and ducted throughout trailer.

B. Tubing for LP gas should be installed from the furnace to the hitch. Recess gas line frame of vehicle. Provide stand and automatic switchover regulator on hitch to take two 100-lb propane bottles. Provide no gas bottles.

C. An air conditioning unit of 1-ton capacity ducted to all three rooms. To operate on same circuit as furnace.

VII. Wiring

A. Provide two grounded, 12-3 size, 20-amp rubber 50 ft drop cords. Two exterior waterproof inlet boxes to be located on right front and left rear of trailer at least 6′ above the trailer floor. Each inlet box to have two twist lock male plugs for 110 VAC service with covers, 20-A size.

B. Grounded wiring throughout the trailer. Two circuits: one for lighting receptacles and fans and power outlets, and the other for furnace and air conditioner.

C. Circuit breakers (not fuses) to be located on wall of the control room.

D. Each of the three rooms to have a covered 200-W fixture, positioned as indicated in Fig. 3.

E. Each fixture must be controlled by a separate dimmer switch, with all three switches being housed in the control room as indicated in Fig. 2.

F. All wiring to meet national wiring standards. Outlets indicated on Fig. 3.

VIII. Interior Fixtures

A. For each partition, a shelf that can be folded down against the partition when not in use. Each shelf: Formica top, 30

inches off floor on partition, in control room. Each shelf should have two sturdy fold-out supports (See Fig. 2).

IX. Miscellaneous

A. Tandem axles with electric brakes and wheel wells that do not intrude into body. Each axle 6000-lb capacity.

B. Four leveling jacks.

C. A heavy-duty hitch with a 2 5/16-inch female socket 18½ inches from the ground with ball.

D. Cable on the hitch for tail, stop, and turn lights, and trailer brakes. Cable connectors not necessary. All of these cables to be attached to the underside of the trailer in such a fashion that children who crawl under the trailer can not rip the cable loose, possibly by recessing wires for all lights into frame of vehicle.

E. Tires 7 × 14.5 inches, 8 ply.

F. All necessary equipment to meet safety lane requirements including safety chains, safety lane inspection, and sticker.

G. Furnish two sets mobile home–type steps with tubular steel frames and wooden treads.

H. Furnish 24 sq ft of interior paneling for future construction of modified panels.

I. Furnish with your bid proposed conditions of standard factory warranty.

J. The company awarded the contract must furnish complete plans for approval and possible slight alterations prior to construction.

K. Completed vehicle to be available for inspection no fewer than 14 days prior to delivery; the company to notify the purchaser when vehicle is ready for inspection.

L. Provide delivery to where vehicle is wanted. (Attach Fig. 2, 3, and 4, or their equivalents).

REFERENCES

BIJOU, S. W. A child study laboratory on wheels. *Child Development*, 1958, 29, 425-427.

BERGMAN, M. O. A mobile laboratory for research in child psychology. Unpublished Masters Thesis, University of Minnesota, 1964.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 153–165 (1969)

Social Reinforcement as a Function of Task Instructions, Sex of *S*, Age of *S*, and Baseline Performance[1]

KENNEDY T. HILL
University of Illinois, Urbana

AND

BARBARA E. MOELY
University of Hawaii

Forty boys and forty girls at each of two age levels (Grades 1 and 2 vs Grades 4 and 5) performed at a marble-dropping task, which was introduced as either a game or a test. Four females served as *Es*. The *Es* were nonresponsive to *S* during a baseline minute, and during the subsequent 5 minutes of the task either made supportive comments (Social Reinforcement Condition) or continued to play a neutral role (Nonreinforcement Condition). Within each age level, *Ss* were blocked according to the sample's median base rate level of response, and base rate was crossed with the other variables in the analysis of difference scores. This procedure minimized group differences in baseline performance, a desirable control in the analysis of difference scores. For older boys, social reinforcement resulted in higher difference scores under game than test instructions, whereas nonreinforcement produced higher difference scores for test than game instructions. For older girls, an opposite pattern of results was obtained, resulting in the significant triple interaction between Sex of *S*, Reinforcement Condition, and Task Instructions for older *Ss*. For younger *Ss*, social reinforcement produced higher difference scores in general than did nonreinforcement. At both age levels, base rate level of *S* did not interact with any of the major findings. Base rate level of response was, however, a highly significant general determinant of change in rate of response, with both older and

[1] This paper is based on the study reported by the authors in a paper read at the annual convention of the American Psychological Association, New York, New York, September 3, 1966. The writers express their gratitude to Sheila Dendurand, Rhonda Fraser, and Corin Nirenstein for serving as experimenters, and to Harold W. Stevenson for his suggestions and assistance in the conduct of this study. The research was supported in part by Grant M-3519 from the National Institute of Mental Health, U. S. Public Health Service; also, by a Public Health Service Fellowship (No. 5 TL MH-6668-06) from the National Institute of Mental Health, U. S. Public Health Service, to the first author, and by a Public Health Service Fellowship (No. HD-01136-02) from the National Institute of Child Health and Human Development, U. S. Public Health Service, to the second author. Author Hill's address: 196 Davenport Hall, University of Illinois, Urbana, Ill., 61801.

younger *Ss* of high base rate level showing steady decreases in performance across the subsequent 5 minutes of the task. The results were related to previous findings and the advantages of blocking on baseline in analyses of change were discussed in general and in light of the present findings.

Recent findings (Hill, 1967; Stevenson and Hill, 1965) have led to the expectation that the effects of social reinforcement and nonreinforcement from adults will be determined in part by subject and task characteristics that involve evaluation of *S*'s competence of performance. Hill (1967), for example, found that the effects of social and nonreinforcement varied according to whether *S* had experienced success or failure in pretraining and whether he admitted to a low or high level of test anxiety as defined by Sarason's Test Anxiety Scale for Children. For low test anxious *Ss* presumably not overly concerned about adults' evaluation of their performance, change in rate of response was higher under social reinforcement after success than after failure, whereas under nonreinforcement performance was higher after failure than success. For high test anxious *Ss*, an opposite pattern of results was obtained, with performance higher under social reinforcement after failure and nonreinforcement after success.

The present study tested for the generality of Hill's (1967) findings by using two different procedures hypothesized as involving evaluation of *S*'s performance. The first procedure was to introduce and emphasize the experimental task as a game or a test situation. This procedure has been used often in studies of anxiety in children (Ruebush, 1963). The other procedure used in the present study was the sex of *S* relative to the sex of *E*, with four female *E*s testing boys and girls. Recent studies including nonreinforcement conditions (Hill and Stevenson, 1965; Stevenson and Knights, 1962) have revealed a "same sex" effect such that an *E* of the same sex as *S* who is nonresponsive as the child performs at a marble dropping task obtains a higher level of performance than does an *E* of the opposite sex of *S*. This effect was attributed by Hill and Stevenson (1965) to a greater anxiety arousal presumed to result from the lack of informative feedback about *S*'s performance from an adult *E* of the same sex as *S* than an *E* of the opposite sex. It appears that as children become older, evaluation from the same-sex parent and same-sex adults in general becomes more important to the child than evaluation from the opposite-sex parent and adults (Stevenson, Hale, Hill, and Moely, 1967).

The present study thus investigated the effects of social reinforcement and nonreinforcement from female adults on boys and girls performing at a marble dropping task introduced as a game or a test. The sample

included younger and older elementary school children. The study tested for whether performance under social and nonreinforcement as a function of game versus test task instructions and whether *S* was of the opposite or of the same sex as *E* would be similar to Hill's (1967) findings for success versus failure in pretraining and whether *S* was low or high test anxious, respectively. Comparable results would be obtained if, for boys, performance for female *E*s under social reinforcement was higher for game than test instructions, while under nonreinforcement performance was higher for test than game instructions; for girls, an exactly opposite pattern of results would be expected, with performance higher under social reinforcement with test instructions but higher under nonreinforcement with game instructions. It was seen as an empirical question as to whether these relations would extend to younger elementary school children, for whom evaluative aspects of the task instructions and the Sex of *S*—Sex of *E* relation may not yet be operative.

The present study also provided an opportunity to test directly for the effects of base-rate level of performance on later performance. In previous studies, the effects of social reinforcement have usually been assessed by a difference score measuring change in rate of response from a baseline period, during which *E* is nonresponsive, to the experimental period during which *E* either delivers social reinforcement or continues to remain nonresponsive (Stevenson, 1965). Parton and Ross (1965) have pointed out that baseline performance may account for group differences in changes in performance if groups differ in base rate level of response. Parton and Ross suggested an analysis, however, which fails to add information to that provided by the separate analyses of base rate and difference scores that are reported in most studies of social reinforcement. In commenting on Parton and Ross's remarks, Stevenson and Hill (1966) suggested that one way to control for base rate effects in analyses of difference scores would be to block on base rate within groups in the experimental design. If base rate could be crossed with all of the other variables under study, group differences in baseline performance would be minimized, which would in turn control for between-group differences in the effects of base rate on change in performance. This procedure would also provide tests for the possible interaction of base rate with the other variables and would account for some of the variance within groups previously treated as error variance, since there is usually a moderate negative correlation between base rate and difference scores. It was found feasible in the present study to cross base rate with the other variables under study with a small reduction (17%) in the experimental sample.

METHOD

Subjects

The initial 192 *Ss* included 24 boys and 24 girls from each of four grade levels: First, second, fourth, and fifth grade. The *Ss* were from an elementary school in Minneapolis, Minnesota, and were of lower-middle socioeconomic class and of average intelligence. The first- and second-grade *Ss* and the fourth- and fifth-grade *Ss* were combined to form the Younger and Older groups of *Ss*, respectively.

Experimenters

There were four female *Es*, including three graduate students in child psychology and a senior in elementary education. The *Es* were 20–24 years of age. An attempt was made to standardize the behavior of the *Es* in order to minimize differences in their administration of the task. The second author served as one *E*. The other three *Es* were not told the purpose of the study until all data had been collected.

Apparatus

The apparatus for the experimental task has been described in detail (Stevenson and Fahel, 1961). It consisted essentially of a table with two bins, 8×10 inches, and a 4-inch transverse upright panel toward the back of the table to shield *E*'s recording from *S*. The bin to *S*'s left contained approximately 1500 marbles of different colors. The bin to *S*'s right was covered with a plate with six $\frac{5}{8}$-inch holes placed randomly about the surface. Below the right bin was a mechanism whereby each marble was counted as it fell to the bottom of the bin. An electrical counter was connected to this mechanism, and placed on the table behind the transverse panel so as to be out of sight of *S*. Timing was done with a stopwatch.

Procedure

Task instructions. The *Ss* were tested individually in one of two vacant rooms in the school building. Each *S* received the marble dropping task under one of two Task Instruction conditions. In the "Game Condition," the task was introduced as a game which some of the other children had already played. The *Ss* were told to play the game by picking up the marbles one at a time, with one hand, and to put any color marble in any hole. The *E* demonstrated the task by dropping six marbles at a rate of approximately one every 2 seconds. The *E* said she would tell *S* when to stop.

In the "Test Condition," *E* referred to the task as a test which some

of the other children had already taken, and which *S* was now going to take. Except for the designation of the task as a test, the instructions were the same as those of the Game Condition.

Experimental task. During the first minute of the marble-sorting task, the base-line period, *E* assumed a neutral role by remaining silent, not looking at *S*, and attending to papers on her side of the transverse panel. During the next 5 minutes, the experimental period, *E* either made a supportive statement about *S*'s performance after *S*'s first response within each 20-second interval (Social Reinforcement condition) or continued to maintain a neutral role throughout the session (Nonreinforcement Condition). The fixed interval schedule used in the Social Reinforcement Condition has the advantage of providing a constant number of supportive statements made to each *S* during the experimental session but the disadvantage of not being contingent upon *S*'s rate of response (Parton and Ross, 1965). Similar effects have been obtained with interval and ratio schedules providing comparable densities of reinforcement in previous studies of social reinforcement (Stevenson and Hill, 1966), and the interval schedule was used in the present study to make the procedure comparable to that of the earlier study by Hill (1967). The six supportive statements used were: "You're doing well, Very good, You're doing fine, That's good, That's fine, and Good." The six statements were arranged in twelve random orders, and two of the random orders and half of a third order were used with each *S*. During the 6 minutes of the task, *E* recorded from the electric counter the number of marbles sorted per minute.

Design

Each of the four *E*s tested 48 children, including three boys and three girls at each age level in each of the four possible task instruction (Game-Test) and reinforcement condition (Social-Nonreinforcement) combinations. The *S*s were randomly assigned to *E*s and to experimental treatments. This procedure resulted in a 4 (Experimenter) $\times$ 2 (Age Level of *S*) $\times$ 2 (Sex of *S*) $\times$ 2 (Task Instructions) $\times$ 2 (Reinforcement Condition) factorial arrangement.

RESULTS

Treatment of Base Rate

Following the suggestion of Stevenson and Hill (1966), it was intended to cross base rate level of *S* with the other variables under study in the analysis of difference scores if such a procedure did not substantially reduce the sample size. A 4 (*E*) $\times$ 2 (Age of *S*) $\times$ 2 (Sex of *S*) $\times$ 2

(Task Instructions) × 2 (Reinforcement Condition) analysis of variance of the base-rate scores revealed that the only significant effect was that of Age of *S* ($F = 140.5$, $df = 1/128$, $p < .001$), reflecting the markedly higher mean base rate of older *S*s (37.9) than younger *S*s (25.6), similar to earlier findings (Stevenson, 1965). Since there were no significant differences among *E*s, either in general or in interaction with other variables, the difference scores were analyzed by collapsing across *E*s and introducing base-rate level of *S* as a factor. Because of the large difference in the mean base rate of younger and older *S*s, base-rate level of *S* could not be crossed with Age of *S* and instead the two age groups were analyzed separately. The *S*s at each age level were divided into two groups on the basis of the median base rate for that age level, 38 for older *S*s, and 25 for younger *S*s. At each age level, it was possible to obtain, within each cell of 12 *S*s, five *S*s above the median and five *S*s below the median base rate. In each cell, two *S*s with scores closest to the median were discarded. By following this procedure, 160 of the original sample of 192 *S*s (83%) could be used in analyses of variance of difference scores in which Base-Rate Level of *S* (High vs Low) was crossed with all of the other variables in the analysis. The procedure resulted in minimal differences in base rate among groups at each age level, except, of course, between *S*s of low versus high base rate.[2] Thus the analyses of variance of the difference scores for both older and younger *S*s involved a 2 (Sex of *S*) × 2 (Task Instructions) × 2 (Reinforcement Condition) × 2 (Base Rate Level of *S*) factorial arrangement with 5 *S*s in each cell.

Difference Scores

The major score used in the analysis of the results was a difference score obtained separately for each *S* by subtracting the number of responses during the first, base-line minute from the number of responses made during each of the 5 subsequent minutes comprising the experimental period. Such a difference score is used in an attempt to gain some control over individual differences in base rate of response (Parton and Ross, 1965, 1967; Stevenson and Hill, 1966), in this case for *S*s at low and high levels of base rate.

Relation to base rate scores. As in previous studies (Stevenson, 1965), difference scores were found to be negatively related to base rate scores within each cell. The median correlations of difference scores from the 5 minutes of the experimental period with base rate scores for the eight

[2] For younger *S*s, the mean base rate for the low base rate group was 19.8, and for the high base rate group, 31.5. For older *S*s, the mean base rate for the low and high base rate groups was 31.2 and 44.3, respectively.

subgroups receiving social reinforcement ranged from —.13 to —.50, with a median of these medians of —.43. For the eight subgroups in the Nonreinforcement Condition, the median correlations ranged from —.23 to —.63, with a median of the medians of —.52. These correlations are similar in strength to those found in previous studies (Parton and Ross, 1965; Stevenson and Hill, 1966).

Older Ss. Difference scores obtained by Older *S*s during the experimental period were subjected to a 2 (Sex of *S*) × 2 (Task Instructions) × 2 (Reinforcement Condition) × 2 (Base Rate Level of *S*) × 5 (Minutes) analysis of variance. It will be recalled that the major effect of interest was the interaction of Sex of *S* × Task Instructions × Reinforcement Condition. This triple interaction was significant ($F = 4.94$, $df = 1/64$, $p < .05$) and reflected a clearly different pattern of results for the experimental variables for boys and girls. As can be seen in Fig. 1, in which the eight means bearing on the triple interaction are presented, for boys performance is similar in the game condition under social and nonreinforcement and at the baseline level of response, whereas for the test instructions performance is higher under nonreinforcement and lower under social reinforcement. For girls, in contrast, performance is higher under nonreinforcement than social reinforcement with the game instructions but higher under social reinforcement and somewhat lower under nonreinforcement with the test instructions. Thus the effects of game vs test instructions on performance under the two reinforcement conditions are opposite for older boys and girls.

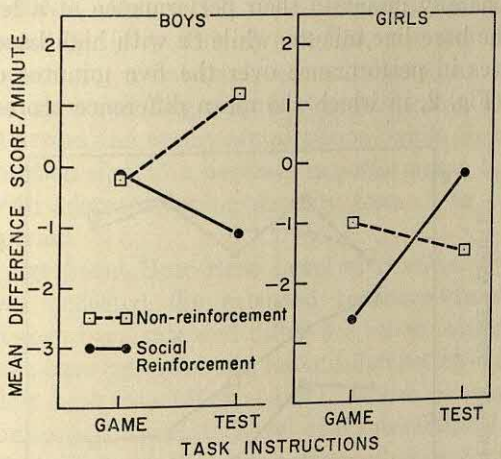


FIG. 1. Mean difference scores, per minute, for older boys and girls tested in each of the four Task Instructions—Reinforcement Conditions combinations.

As can also be seen in Fig. 1, the significant main effect of Sex of *S*

($F = 4.50$, $df = 1/64$, $p < .05$) reflects the generally higher difference scores obtained for boys than girls; boys show little change in rate of response following the baseline minute ($M = -.08$), whereas girls show a general decrease in rate of response following the baseline minute ($M = -1.34$).

The effect of Base-Rate Level of *S* on difference scores was highly significant ($F = 43.81$, $df = 1/64$, $p < .001$). As would be expected from the negative correlation between base rate and difference scores obtained within all subgroups, *Ss* with high base rates obtained a lower mean difference score ($M = -2.67$) than *Ss* with low base rates ($M = +1.24$). In addition, the significant interaction of Minutes × Base Rate Level of *S* ($F = 8.87$, $df = 4/256$, $p < .001$) reflects the tendency for *Ss* with low base rates to generally maintain their performance at a level somewhat above that of the base-line minute, while *Ss* with high base rates tended to show decreases in performance over the five minutes of the experimental period (Fig. 2, in which the mean difference scores, per minute,

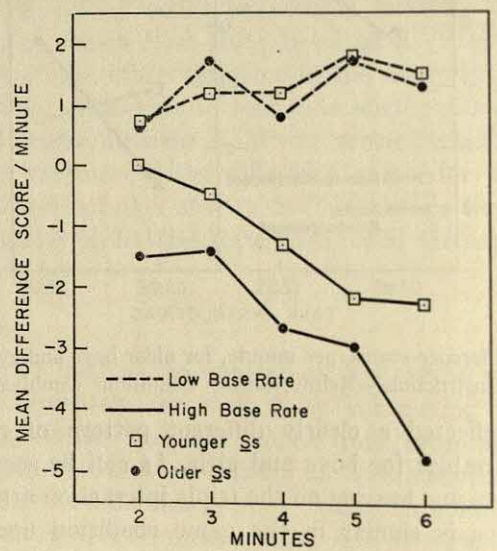


FIG. 2. Mean difference scores, per minute, across the experimental period for older and younger *Ss* of low and high base rate level.

are shown for older and younger *Ss* of low and high base-rate level). This latter pattern was reflected in a significant Minutes effect ($F = 7.52$, $df = 4/256$, $p < .001$), with difference scores decreasing in general across the 5 minutes of the experimental period, due to the marked decreases for the high base rate *Ss*. There were no other significant effects for the older *Ss*.

An analysis of the between-subjects effects involving base-rate level of response revealed that including base rate as a crossed variable reduced the error term by 38% from the case in which base-rate level was excluded from the analysis. The corresponding reduction of error in the within-subjects portion of the analysis of variance due to the addition of base rate as a crossed variable was 8%.

Younger Ss. The difference scores of younger *Ss* were also subjected to a 2 (Sex of *S*) $\times$ 2 (Task Instructions) $\times$ 2 (Reinforcement Condition) $\times$ 2 (Base-Rate Level of *S*) $\times$ 5 (Minutes) analysis of variance. The significant Reinforcement Condition effect ($F = 5.32$, $df = 1/64$, $p < .05$) reflected the fact that *Ss* receiving social reinforcement obtained higher difference scores in general ($M = +.75$) than *Ss* in the Nonreinforcement Condition ($M = -.72$). The Minutes $\times$ Social Reinforcement Condition interaction ($F = 4.80$, $df = 4/256$, $p < .01$) indicates, moreover, that *Ss* receiving social reinforcement showed an increase in performance for each succeeding minute, with difference scores increasing from .4 to 1.0 across the experimental period, while *Ss* in the Nonreinforcement Condition showed a decrease in performance for each succeeding minute, with scores dropping steadily from .5 to -1.9 across the experimental period.

The highly significant Base-Rate Level of *S* effect ($F = 15.51$, $df = 1/64$, $p < .001$) reflected the expected tendency from the negative correlation between base rate and difference scores within all subgroups for *Ss* with a high base rate to obtain lower difference scores ($M = -1.24$) than *Ss* with low base rate ($M = +1.27$). As can be seen in Fig. 2, the significant Minutes $\times$ Base-Rate Level of *S* interaction ($F = 5.62$, $df = 4/256$, $p < .001$) was due to the same pattern of scores found for older *Ss*: For the younger *Ss*, *Ss* with low base rates showed a small increase in performance over the experimental period above that of baseline performance, while *Ss* with high base rates showed a steady decrease in performance from base line across the experimental period.

The only other significant effect in the analysis for younger *Ss* was the Minutes $\times$ Base-Rate Level of *S* $\times$ Task Instructions triple interaction ($F = 3.26$, $df = 4/256$, $p < .05$). For *Ss* with low base rates, difference scores tended to increase across the experimental minutes for *Ss* in the Test Condition (from .20 to 2.15 per minute), but to decrease slightly for *Ss* in the Game Condition (from 1.20 to .80); for *Ss* with high base rates, performance tended to decrease more over the experimental period for *Ss* in the Test Condition (from 1.15 to -2.45) than for *Ss* receiving Game Instructions (from -1.10 to -2.10). For younger children, then, *Ss* in the Test Condition tended to contribute more to the Minutes $\times$ Base Rate interaction effect than *Ss* in the Game Condition.

An analysis of the between subjects effects for younger *Ss* revealed that including base rate as a crossed variable reduced the error term by 16% from the case in which base-rate level was excluded from the analysis. The corresponding reduction of error in the within-subjects portion of the analysis of variance due to the addition of base rate as a crossed variable was 6%.

DISCUSSION

The results for the older *Ss* indicate an affirmative answer to the question of whether the effects on performance of game versus test instructions and the sex of *S* relative to the sex of *E* are similar to the effects of success-failure experiences in pretraining and low versus high test anxiety of *S*. The different pattern of results for boys and girls under the two task instructions in the present study are very similar to the pattern Hill (1967) obtained for low and high test anxious *Ss* under the two pretraining conditions. The basis for these similar findings, however, is not entirely clear. Commonality in the effects of the task instructions, pretraining experiences, and test anxiety level of *S* might be expected from the general use of these three variables to manipulate children's anxiety in evaluative situations (Ruebush, 1963). Test instructions, failure in pretraining, and a high level of test anxiety of *S* would all be expected to heighten a child's awareness of the potential for failure or poor performance at a task and of possible disapproval from the adult *E*. Why the sex of *S* relative to the sex of *E* should have similar effects as these other three variables has not been well established. The performance data of Hill and Stevenson (1965) and the preferences for performing for adults seen on film obtained by Stevenson *et al.* (1967) offer suggestive evidence that evaluation from same-sex adults is more important and more likely to elicit anxiety in children than evaluation from opposite-sex adults. The present results are consistent with this notion in that the performance of boys and girls tested by female *Es* (as a function of task instructions) is similar to the performance of low and high test anxious *Ss* obtained by Hill (1967) (as a function of pretraining experiences). These results taken together, however, are far from definitive. The determinants of boys and girls differential reactions to male and female adults are undoubtedly quite complex and determined by a number of interrelated factors (Stevenson, 1965). Until the sex of *S* relative to the sex of *E* is related to independent and more direct measures of anxiety involving evaluation and behavioral correlates of such anxiety, the hypothesis that evaluation from same-sex adults arouses greater anxiety of this nature is concluded to be tenable but unconfirmed.

An important difference in the results of the present study and the earlier investigation by Hill (1967) for the variables under discussion is that an overall sex difference was obtained in the present study but a general test anxiety level of *S* effect was not obtained in the earlier study. This finding also argues against the conclusion that the effects of sex of *S* relative to the sex of *E* and test anxiety are directly analogous. Examination of results of earlier studies reveals that the tendency found in the present study for boys to show small changes in performance but for girls to show general decrements in performance across the experimental period is significant in some studies (Hill and Stevenson, 1964) and often appears as a trend in other studies (Hill, 1967, as reported in Hill, 1965; Stevenson, 1961; Stevenson and Hill, 1965). Whether the effect is due to sex differences concerning the intrinsic interest of the task, the manner in which *S*s interpret the task, the degree to which *S*s are distracted from the task by *E*s presence or verbal comments to *S*, or some other factor(s) remains a question for future research.

The performance of younger *S*s in the present study was controlled by whether or not the experimenter was making supportive comments. Under social reinforcement younger *S*s performed at a somewhat higher level across the experimental period, while under nonreinforcement performance fell off steadily. It appears that the effects of the evaluative properties of the task instructions and the sex of the child relative to the sex of the adult are not yet operative for younger elementary school age children. Allen (1966) has also found that younger children's persistence on simple tasks seems to be strongly determined by whether or not the adult *E* is supportive or not.

The strong and consistent base-rate results have bearing on recent evaluations of the methodologies used in studies of social reinforcement with children (Parton and Ross, 1965, 1967; Stevenson and Hill, 1966). Parton and Ross (1967, p. 324) have recently concluded from this discussion that:

> Stevenson and Hill now recommend the use of a factorial design in which levels of base rate are entered as a factor. Aware that the analysis problem arises when the experimental groups differ in terms of base-rate mean and/or variance, Stevenson and Hill suggested that levels of base rate be nested within each experimental group. Unfortunately, the interaction between a nested factor and the factor under which it is nested cannot be evaluated. Hence, this design will not serve the function suggested by Stevenson and Hill.

These comments by Parton and Ross are misleading. Stevenson and Hill (1966, pp. 324–325) noted that base rate could be crossed with experimental variables but would have to be nested within subject variables which showed a difference in base rate. The present study

demonstrates the feasibility of crossing base rate with all of the variables studied except age of *S* with a relatively small (17%) sample loss. The separate analyses of variance for each age level demonstrated all three advantages of blocking on base rate suggested by Stevenson and Hill. First, and most critical, the procedure minimized base-rate differences among subject and experimental groups by assuring an approximately equal mean and distribution of base-rate scores in each group. Thus base rate differences are not confounded with the effects of subject and experimental variables, a confounding which, as Parton and Ross (1965) point out, is a major problem in the use of difference scores. Second, the procedure made it possible to assess for, and in the present study rule out, interactions of other variables with base-rate level of *S* which might compromise conclusions concerning the effects of these other variables of primary interest. Parton and Ross (1967) are correct that such interactions can not be evaluated for variables under which base rate has to be nested. Third, blocking on base rate in order to control for baseline differences had the desirable effect of reducing error variance by as much as 38% by accounting for within group variance due to the negative correlation between base-rate and difference scores. The tendency for *Ss* with a low base rate to show stable performance but for *Ss* with a high base rate to show marked decreases in performance across the experimental period (Fig. 2) is likely to be a result of fatigue operating more strongly for the latter group of *Ss*.

It should be possible in future studies to randomly assign an equal number of *Ss* of low and high base rate to each experimental condition with only a very small sample loss in filling the last few cells, since the median base rate for *Ss* of a given age is fairly stable across samples (Stevenson, 1965). As Stevenson and Hill (1966) suggest, it may be more difficult to cross base rate with subject variables (and pretraining variables), although at present, age is the only individual difference (or pretraining) variable that has been consistently and strongly related to base rate of responding (Stevenson, 1965). Whether or not the necessity of employing a wholly or partially nested design is a serious difficulty in a study will depend on the design and goals of the particular study. In the present study, for example, base-rate level of *S* could not be compared directly across age, since marked age differences were obtained for base-rate scores. This did not cause serious difficulty in the analysis of difference scores, however, since interest was primarily in the pattern of effects appearing within each age level.

It is suggested that base rate be entered as a factor in future studies of social reinforcement whenever it is feasible to do so, and that base rate be crossed with other variables under study if such a procedure does

not result in serious problems, e.g. a large sample loss or a nonrepresentative sample. Such blocking procedures might be useful, moreover, in any research in which base-level performance is related to later changes in performance (Harris, 1963), i.e. whenever determinants of base line are determinants of change.

REFERENCES

Allen, S. The effects of verbal reinforcement on children's performance as a function of type of task. *Journal of Experimental Child Psychology*, 1966, **3**, 57–73.

Harris, C. W. (Ed.) *Problems in measuring change*. Madison: University of Wisconsin Press, 1963.

Hill, K. T. The effects of social reinforcement and nonreinforcement following success and failure as a function of test anxiety of *S*, the sex of *S*, and the sex of *E*. Unpublished doctoral dissertation, University of Minnesota, 1965. University Microfilm Number: 65–15, 265.

Hill, K. T. Social reinforcement as a function of test anxiety and success-failure experiences. *Child Development*, 1967, **38**, 723–737.

Hill, K. T. and Stevenson, H. W. Effectiveness of social reinforcement following social and sensory deprivation. *Journal of Abnormal and Social Psychology*, 1964, **68**, 579–584.

Hill, K. T. and Stevenson, H. W. The effects of social reinforcement vs. nonreinforcement and sex of *E* on the performance of adolescent girls. *Journal of Personality*, 1965, **33**, 30–36.

Parton, D. A. and Ross, A. O. Social reinforcement of children's motor behavior: a review. *Psychological Bulletin*, 1965, **64**, 65–73.

Parton, D. A. and Ross, A. O. A reply to 'the use of rate as a measure of response in studies of social reinforcement.' *Psychological Bulletin*, 1967, **67**, 323–325.

Ruebush, B. K. Anxiety. In H. W. Stevenson (Ed.), *Sixty-second yearbook of the national society for the study of education, part I: Child Psychology*. Chicago: University of Chicago Press, 1963, 460–517.

Stevenson, H. W. Social reinforcement of children's behavior. In L. P. Lipsitt and C. C. Spiker (Eds.), *Advances in Child Development and Behavior*. Vol. II, New York: Academic Press, 1965, 97–126.

Stevenson, H. W. Social reinforcement with children as a function of CA, sex of *E*, and sex of *S*. *Journal of Abnormal and Social Psychology*, 1961, **63**, 147–154.

Stevenson, H. W. and Fahel, L. S. The effect of social reinforcement on the performance of institutionalized and noninstitutionalized normal and feebleminded children. *Journal of Personality*, 1961, **29**, 136–147.

Stevenson, H. W., Hale, G. A., Hill, K. T. and Moely, B. E. Determinants of children's preferences for adults. *Child Development*, 1967, **38**, 1–14.

Stevenson, H. W. and Hill, K. T. The effects of social reinforcement and nonreinforcement following success and failure. *Journal of Personality*, 1965, **33**, 418–427.

Stevenson, H. W. and Hill, K. T. The use of rate as a measure of response in studies of social reinforcement. *Psychological Bulletin*, 1966, **66**, 321–326.

Stevenson, H. W. and Knights, R. M. The effectiveness of social reinforcement after brief and extended institutionalization. *American Journal of Mental Deficiency*, 1962, **66**, 589–594.

Statement of ownership, management and circulation required by the Act of October 23, 1962, Section 4369, Title 39, United States Code; of

Journal of Experimental Child Psychology

Published quarterly at Exchange Place, Hanover, Pa. 17331 for Academic Press Inc., 111 Fifth Avenue, New York, N. Y. 10003. Editor: Sidney W. Bijou, University of Illinois, Champaign, Illinois 61820. Owned by Walter J. Johnson and Thekla Johnson, 111 Fifth Avenue, New York, N. Y. 10003. Known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages or other securities: None.

Paragraphs 2 and 3 include, in cases where the stockholder or security holder appears upon the books of the company as trustee or in any other fiduciary relation, the name of the person or corporation for whom such trustee is acting, also the statements in the two paragraphs show the affiant's full knowledge and belief as to the circumstances and conditions under which stockholders and security holders who do not appear upon the books of the company as trustees, hold stock and securities in a capacity other than that of a bona fide owner. Names and addresses of individuals who are stockholders of a corporation which itself is a stockholder or holder of bonds, mortgages or other securities of the publishing corporation have been included in paragraphs 2 and 3 when the interests of such individuals are equivalent to 1 percent or more of the total amount of the stock or securities of the publishing corporation.

Total no. copies printed, average no. copies each issue during preceding 12 months: 1400; single issue nearest to filing date: 1500. Paid circulation (a) to term subscribers by mail, carrier delivery or by other means, average no. copies each issue during preceding 12 months: 1151; single issue nearest to filing date: 1262. (b) Sales through agents, news dealers, or otherwise, average no. copies each issue during preceding 12 months: 0; single issue nearest to filing date: 0. Free distribution by mail, carrier delivery, or by other means, average no. copies each issue during preceding 12 months: 74; single issue nearest to filing date: 81. Total no. of copies distributed, average no. copies each issue during preceding 12 months: 1225; single issue nearest to filing date: 1343.

(Signed) J. B. Chertok, Vice President

Journal of
APPLIED BEHAVIOR ANALYSIS

MONTROSE M. WOLF, Editor
DONALD M. BAER, Associate Editor
VICTOR G. LATIES, Executive Editor

EDITORIAL BOARD

T. Ayllon
N. H. Azrin
A. Bandura
W. C. Becker
J. S. Birnbrauer
C. B. Ferster
I. Goldiamond
J. G. Holland
F. S. Keller
P. Lang
O. R. Lindsley
O. I. Lovaas
J. L. Michael
G. R. Patterson
T. R. Risley
J. A. Sherman
M. Sidman
G. Siegel
B. F. Skinner
J. E. Spradlin
A. W. Staats

Second quarterly issue, SUMMER, 1968

Behavioral engineering: postural control by a portable operant apparatus.

N. Azrin, H. Rubin, F. O'Brien, T. Ayllon, and D. Roll.

Establishing use of descriptive adjectives in the spontaneous speech of disadvantaged preschool children.

Betty M. Hart and Todd R. Risley.

Effects of candy and social reinforcement, instructions, and reinforcement schedule leaning on the modification and maintenance of smiling.

B. L. Hopkins.

Feedback in behavior modification: an experimental analysis in two phobic cases.

Harold Leitenberg, W. Stewart Agras, Laurence E. Thomas, and Dale E. Wright.

Rules, praise, and ignoring: elements of elementary classroom control.

Charles H. Madsen, Jr., Wesley C. Becker, and Don R. Thomas.

An experimental analysis of verbal imitation in preschool children.

Thomas A. Brigham and James A. Sherman.

Operant conditioning of eye movements.

Stephen R. Schroeder and J. G. Holland.

Collateral social development accompanying reinforcement of outdoor play in a preschool child.

Joan Buell, Patricia Stoddard, Florence Harris, and Donald M. Baer.

A method to integrate descriptive and experimental field studies at the level of data and empirical concepts.

Sidney W. Bijou, Robert F. Peterson, and Marion H. Ault.

Address subscription orders to Mary L. Sherman,
Department of Human Development, University of Kansas
Lawrence, Kansas 66044

Students $4 * Individuals (personal use) $8 * Institutions $16

The Columbus

Picture Analysis of Growth Towards Maturity
A Series of 24 Pictures and a Manual

M.J. Langeveld (Utrecht)
Translated by G. Uildriks, Utrecht

IV + 72 p., 24 pictures (of which 3 in colour), 3 tab., 1969.
S.Frs./DM 59.–/US $ 14.15/118s.

Preface and Introduction – Chapter I: The Cards – Chapter II: Subdivision and Use of the Series – Chapter III: Some Remarks Concerning Projection, Maturity, Emancipation – References –

'The Columbus' consists of 24 pictures of which 3 coloured. It is designed for children and young people from 5 to 18 years old and it has been developed with a 20 years' follow-up in different social and ethnic groups of western and non-western countries. It is an instrument for the study and analysis of the patterns of emancipation from infantile forms of life and relationship towards mature relations to oneself, other people, educators, etc. and towards personal as well as social maturity. Its use is indicated for child guidance clinics, child psychiatrists, child psychologists, psychotherapists, clinical psychologists, etc.

S.Karger Basel (Switzerland) New York

K 102

INFORMATION FOR AUTHORS

The *Journal of Experimental Child Psychology* will publish papers in which the behavior and development of children is clearly related to its determining variables. Typically (but not always) this would mean that a variable has been manipulated in an experimental manner. Articles will be reviewed in terms of the apparent reliability of their findings and the clarity of the relationships they affirm. Research conceived in any theoretical framework is invited, so long as its conclusions are reliable and compelling. Occasionally, critical reviews of related bodies of experimental literature, theoretical discussions which define new paths of experimental work, and well-controlled developmental studies on infra-human subjects will be published. The criterion for acceptance of such papers will be their value in prompting experimental research in related areas with children. Short notes describing new apparatus or techniques, new materials, circuits, or other information valuable to investigators in the field will be accepted from time to time. There is no page charge for publication.

Two copies of each manuscript and each figure or table should be sent to:

Dr. Sidney W. Bijou, Child Behavior Laboratory, 403 E. Healey, University of Illinois, Champaign, Illinois 61820

A third copy should be retained by the author. Manuscripts must be double-spaced and may be typed, dittoed, or mimeographed. Each manuscript must include a concise abstract, preferably not exceeding 300 words, which will appear after the title and author's name in the published article. In some papers this abstract will obviate the need for a summary. In general, authors are free to use whatever organization their paper seems to require, but the specifications of the *Publication Manual* of the American Psychological Association are strongly recommended.

Figures and Tables. Figures should be numbered consecutively with Arabic numerals in order of mention in the text; each figure should have a descriptive legend. Legends should be typed together on a separate sheet, double-spaced. All illustrations should be in finished form ready for the engraver. A convenient size for drawings is 8½ × 11 inches (21 × 27.5 cm). Drawings should be made with India ink on tracing linen, smooth surface white paper, or bristol board. Graph paper if used should not be ruled in blue. Grid lines that are to show in the final engraving should be inked in black. Sharp, clear glossy prints of figures are acceptable.

Tables should be typed on separate pages, numbered consecutively with Arabic numerals in order of mention in the text. All tables should have descriptive headings, typed (double-spaced) above the table.

Literature references in the text should be cited by author's surname and the year of publication, e.g., Smith, 1960; Smith *et al.*, 1959; Smith and Jones, 1961, p. 250 (for references to a specific page). If a reference has more than two authors, the citation includes the surnames of all authors at the first mention, but later citations of the same reference include only the surname of the first author and the abbreviations "*et al.*" Suffixes *a*, *b*, etc., should be used following the date to distinguish two or more works by the same author(s) in the same year, e.g., Smith, 1960a, 1960b. Literature citations in the bibliography should be arranged alphabetically according to the surname of the author.

Journal titles should be written out in full according to the form followed in the Psychological Abstracts (Vol. 39).

The desired style of capitalization and punctuation in references is apparent in the following examples:

Staats, A. W., and Staats, Carolyn K. A comparison of the development of speech and reading behavior with implications for research. *Child Development*, 1962, 33, 831–846.

Spiker, C. C. Research methods in children's learning. In P. H. Mussen (Ed.), *Handbook for research methods in child development*. New York: Wiley, 1960. Pp. 374–420.

Munn, N. L. *The evolution and growth of human behavior*. Cambridge: Riverside Press, 1955.

Reprints: Fifty gratis reprints are supplied and additional copies may be ordered.

Research in Verbal Behavior and Some Neurophysiological Implications

Edited by **Kurt Salzinger**
and **Suzanne Salzinger**
Biometrics Research
New York State Department of Mental Hygiene
and Polytechnic Institute of Brooklyn
New York, New York

This volume contains a complete discussion of the current directions of research in language, including verbal behavior in the light of the principles of neurophysiology. It presents a study of language in man and animals, in its normal state and under the influence of drugs and pathology (aphasia, schizophrenia) in adults and children, based upon models from the areas of learning theory, linguistics, mathematics, psychophysics and neurophysiology. It includes a discussion of problems emerging when techniques derived from animal experimentation are applied to human verbal behavior.

The work covers a wide range of topics, from vocal behavior in animals to emotions and images in man. Fields discussed range from physiology to bilingualism. Disciplines from linguistics and psychology to medicine, psychiatry, psychopharmacology and neurology are represented.

The book will be particularly useful to graduate students and advanced undergraduates taking courses in language as well as to research workers in the area.

CONTRIBUTORS: Marianne Amarel. Louis Cassotta. Richard Allen Chase. Frances E. Cheek. Charles Clifton, Jr. Alberto DiMascio. Stanley Feldstein. Samuel Fillenbaum. Norman Geschwind. Norman Ginsburg. Murray Glanzer. Frieda Goldman-Eisler. Louis A. Gottschalk. Joseph H. Grosslight. Katherine Harris. Gilbert Honigfeld. Davis Howes. Joseph Jaffe. Lyle V. Jones. Leonard Krasner. Wallace E. Lambert. Harlan L. Lane. John C. Lilly. A. R. Luria. Paula Menyuk. Irwin Pollack. Malcolm S. Preston. Kurt Salzinger. Arthur W. Staats. John A. Starkweather. Irene E. Waskow. Joseph M. Wepman. Wesley C. Zaynor.

1967, 510 pp., $11.00

AP 2033

Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

JOURNAL OF Experimental Child Psychology

EDITOR Sidney W. Bijou

ASSOCIATE EDITORS: Richard S. Bogartz
David Zeaman

EDITORIAL BOARD:

Donald M. Baer
Albert Bandura
Wesley C. Becker
Harry Beilin
Gordon N. Cantor
Peter Eimas
Robert L. Fantz
Eleanor J. Gibson
Eugene S. Gollin
Willard W. Hartup
Raymond H. Hohle
Herbert Kaye
Tracy S. Kendler
Helen L. Koch
Eugene R. Long
David S. Palermo
Herbert L. Pick
Hayne W. Reese
Bryan E. Shepp
Murray Sidman
Einar Siqueland
Patrick Suppes
Joachim Wohlwill

Volume 7, Number 2, April 1969

ACADEMIC PRESS
New York and London

CONTENTS

Rochel Gelman. Conservation Acquisition: A Problem of Learning to Attend to Relevant Attributes 167

Bonnie J. Lubker. Setting Similarity and Successive Discrimination Learning by Children 188

Ioannis Paraskevopoulos. Effects of Redundancy on Information-Reduction Tasks 195

Sally A. Sibley, Martha S. Abbott, and Betty P. Cooper. Modification of the Classroom Behavior of a Disadvantaged Kindergarten Boy by Social Reinforcement and Isolation 203

Dorothy Uhr. Spelling Performance of Left-Handed Schoolchildren as Affected by the Use of a Pencil Modified to Increase Visual Feedback . . . 220

J. Allan Cheyne and Richard H. Walters. Intensity of Punishment, Timing of Punishment, and Cognitive Structure as Determinants of Response Inhibition 231

Carolyn Kent Rovee. Psychophysical Scaling of Olfactory Response to the Aliphatic Alcohols in Human Neonates 245

Nancy A. Myers and Jerome L. Myers. A Developmental Study of Sequence Structure in Binary Prediction 255

Ross Ginsberg. Number and Color Responses in the Young Child 265

Jerome Smith and Jeffrey Tunick. Cross-modal Transfer of a Discrimination by Retarded Children 274

Judith Dubignon and Dugal Campbell. Sucking in the Newborn during a Feed 282

Robert H. Wahlsuchby. The Effects of Time-Out from Positive Reinforcement on the Operant Behavior of Preschool Children 299

Grover J. Whitehurst. Discrimination Learning in Children as a Function of Reinforcement Condition, Task Complexity, and Chronological Age . 314

William J. Meyer and Arthur Orgel. The Effects of Age, Stimulus Intensity, and Training Trials on Mediated Stimulus Generalization 326

Bernard Z. Kamsel. The Effect of Age, Complexity, and Amount of Contour on Pattern Preferences in Human Infants 339

Thomas F. Cunningham. Dimensional Transfer and Cue Retention in Retardate Discrimination Learning 355

Judith L. Rapoport. Size-Constancy in Children Measured by a Functional Size-Discrimination Task 366

Merrill M. Mitler and Lauren Harris. Dimension Preference and Performance on a Series of Concept Identification Tasks in Kindergarten, First-Grade, and Third-Grade Children 374

Errata 385

Published bimonthly at Exchange Place, Hanover, Pa. 17331 by Academic Press, Inc.
111 Fifth Avenue, New York, N. Y. 10003.
In 1969, Volumes 7 and 8 (consisting of 3 issues each) will be published.
Price: $22.00
Private subscriptions (for subscriber's personal use): $10.00
All correspondence and subscription orders should be sent to the office of the Publishers at 111 Fifth Ave., New York, N. Y. 10003
Second-class postage paid at Hanover, Pennsylvania 17331

© 1969 by Academic Press, Inc.

Additional Volume of *Journal of Experimental Child Psychology* to be published in 1969

In order to permit more rapid publication of a number of important manuscripts now on hand, an additional volume will be released during 1969.

The revised publication schedule is as follows:

Volume 7—February, April, June
Volume 8—August, October, December

The price of each 1969 volume is $22.00 for institutional subscribers; and $10.00 for private subscribers certifying that the journal is for personal use.

Academic Press, Inc.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 167–187 (1969)

Conservation Acquisition: A Problem of Learning to Attend to Relevant Attributes[1]

ROCHEL GELMAN[2]

Brown University

Five-year-old children who failed on conservation tests of length, number, mass, and liquid amount were given discrimination learning set (LS) training on length and number tasks. Posttests of conservation showed near perfect specific (length and number), and approximately 60% nonspecific (mass and liquid amount) transfer of training. This effect was durable as measured 2–3 weeks later. Analyses of LS learning results and the effects of other training conditions support the hypothesis that young children fail to conserve because of inattention to relevant quantitative relationships and attention to irrelevant features in classical conservation tests.

In general, a test for a child's ability to conserve quantity involves the following sequence of events: (1) An *S* is shown two identical objects or sets of objects; (2) he is then asked to judge whether the two objects are quantitatively equal; (3) if *S* says that they are equal, *E* alters some perceptual but no quantitative properties of one of the stimuli; (4) *S* is asked once more if the two objects (changed versus altered) are still equal with respect to amount; (5) and finally *S* is asked to explain his judgement. If *S* says the stimuli still have equal amounts and is able to explain his answer logically (Piaget, 1952), he is judged a "conserver." Alternatively, if he fails to indicate that the amounts are equal or gives a nonlogical explanation, he is judged a "nonconserver."

This general procedure has been used to test for the conservation of a variety of quantitative concepts: number, length, mass, and liquid

[1] Based on a dissertation submitted in partial fulfillment of the requirements for a Ph.D., University of California, Los Angeles. The research was supported by a University of California, Los Angeles Fellowship to the author and by grant MH-08741 to Tom Trabasso from the Institute of Mental Health, USPHS. I thank M. Friedman, W. Jeffrey, E. Keislar, T. Trabasso, and M. Wittrock for serving as members of my dissertation committee; Ludwig Mosberg for serving as a rater; and the staff at Sherman Oaks Elementary School, Sherman Oaks, California, for freely providing research facilities and making their students available for this study. Special thanks are due Tom Trabasso for his encouragement and advice during all phases of this research.

[2] Now at the University of Pennsylvania.

© 1969 by Academic Press, Inc.

amount (Flavell, 1963). The findings from these tests tend to indicate that young children under seven years of age do not conserve, children approximately seven years of age conserve on some tasks, and children eleven years of age and older conserve on all tasks (Inhelder and Piaget, 1958).

Theoretical accounts for these results may be found in several sources (Almy, Chittenden, and Miller, 1966; Bruner *et al.*, 1966; Flavell, 1963; Wallach, 1963; Wallach, Wall, and Anderson, 1967). In many of these explanations, there appears to be an emphasis upon the nature of the internal cognitive structures of the child. Thus for example, there is consideration of the extent to which a child uses one or more logical operations of multiplication, addition and subtraction, compensation, and inversion. Other explanations deal with the effects of misleading cues, set, reinforcement, and conflict. In particular, explicit reference to factors controlling or directing *S*'s *attention* to quantitative attributes are lacking. The present study examined the possibility that a young child's failure on conservation tests may be a function of inattention to the relevant quantitative attributes of the test or attention to irrelevant features such as changes in size, shape, and color. The single important implication of this hypothesis is that a young child may in some way be able to conserve, and would do so were it not for his strong tendencies to attend to stimulus changes (Bruner *et al.*, 1966). This is illustrated by considering the nature of conservation tests and how a young child's attention might operate during such tests.

To begin an assessment of the role of attention in conservation tests, one may first note Zimiles' (1966) observations that a young child variously responds to numerical quantity on the basis of cues such as length, shape, spacing, and actual number, i.e., his definition of number is multidimensional. Of interest is that Trabasso and Bower (1968) have shown that an *S* presented with a multidimensional stimulus responds to the attribute or attributes which are most salient, or attract his attention. If one assumes that, in general, young children do define quantity multidimensionally, then it is possible to show that differential attention to the various "quantity" attributes may determine whether or not a child will behave as a conserver.

To make matters specific, consider the liquid conservation test. To begin, *S* is shown two glasses of water that are the same size, shape, height, and width; and equal amounts of water are poured into each glass. The stimulus complex of each glass with water can be thought of as a multidimensional pattern with at least six attributes; these being size, shape, height, width, water level, and actual amount of water. The *S* may perceive relations and aspects other than those defined by *E*;

hence we shall use the term "cue" to refer to any stimulus attribute to which *S* may attend, code, and use as a basis for his response (Lawrence, 1963). From the *E*'s point of view only one cue is relevant, i.e., related to the solution of the conservation problem, and this is amount of water. All others are irrelevant. However, from the viewpoint of a young *S* all cues are potentially relevant to his definition of amount. When he is asked to judge amount, he may do so on the basis of any or all cues in the complex. At the start of the conservation problem *E* has no way of knowing to which cue *S* is attending. He is, however, most likely attending to one of the *n* irrelevant dimensions rather than the one quantity dimension. Furthermore, when *E* changes a stimulus array, *S*'s attention may be drawn to the irrelevant cues since these all change while quantity does not. In fact, this manipulation should serve to enhance the likelihood of *S* using irrelevant features since movement or change is a way of bringing attention to an attribute. If *S* does use an irrelevant cue at the start of the conservation task, it does not matter, for he will still be able to judge the stimuli as equal. Furthermore, he is not asked to explain his response. What does matter, is which cue *S* attends to after *E* transforms one of the stimuli. If *S* attends to an irrelevant cue, he will judge the amounts as different and therefore not conserve. Since our analysis shows that he is most likely to attend to irrelevant cues, it would seem the conservation task needs be modified to control for attention. Alternatively, *S* could be trained to attend to relevant and ignore irrelevant cues before being tested on the classical conservation problems.

Now consider a comparison between conservation and discrimination learning tests. In both, *S* is shown multidimensional stimulus patterns; presumably can respond to any of the definable aspects therein; and is to attend, and then respond to a relevant attribute (s). Unlike in the conservation test, *feedback* in discrimination training serves to inform *S* as to which attribute he should key his responses, since reward is consistently associated with the relevant, but not the irrelevant attributes. It is always possible for *S* to correct himself. In conservation tests, correction is not possible, since *S* is not told that his answers are wrong nor given chances to try alternative cues. The implication of this comparison is that *S*s might be brought to attend to quantity attributes, and thereby conserve, via discrimination training on problems related to conservation. In this regard, learning set (LS) training procedures seem most appropriate (Braine, 1962). Here *S* undergoes training with a large number of problems containing many different stimuli; although stimuli differ, across problems there is one common relationship, and attentional responses to this common cue are reinforced. The

oddity task is an example. Here *S* is presented with a series of problems and in each, three stimuli are shown—two are identical, the third different. In each case, *S* is rewarded for choosing the "odd" object; concrete features are changing and irrelevant. Subjects learn these problems by ignoring irrelevant hypotheses, or eliminating "error factors" until they come to attend and respond correctly to the relevant aspects (Harlow, 1959).

The above considerations lead us to use oddity LS training to "teach" *Ss* to attend and respond to *quantitative relations*. If *Ss* learn to attend, and respond to quantity and not other relational cues, then they should conserve when transferred to standard conservations tests.

METHOD

General Design

The experiment consisted of three phases: pretesting, training, and posttesting. In pretesting *Ss* were given standard length, number, mass, and liquid amount conservation tests. If *S* failed to conserve (see criterion below), he underwent the remaining two phases, each taking 2 days. Training was initiated within 2 weeks following testing. The training and initial transfer testing were always conducted on 3 consecutive days. A second posttest followed within 2–3 weeks. Each *S* was seen individually.

There were three kinds of training; (1) modified learning set with length and number stimuli; (2) experience with the latter problems without feedback; and (3) modified learning set with "junk" stimuli. Conditions (2) and (3) served as controls for the effects of feedback on the learning of length and number concepts, and general discrimination and labelling experience, respectively. Following training, all *Ss* were retested on length, number, mass, and liquid amount conservation tasks thereby providing tests of specific and nonspecific transfer.

Subjects

The *Ss* were 110 children (57 girls and 53 boys) haphazardly selected from the kindergartens of Sherman Oaks Elementary School in Sherman Oaks, California. Their ages ranged from 4 years, 9 months to 6 years (median age, 5 years 4 months). Three *Ss* were dropped from the experiment because they were unable to count, leaving an N of 107. Of these, 60 were assigned as described below to one of the three training conditions. The median age of the LS group was 5 years 4 ½ months, and for the other two groups, 5 years 5 months.

Conservation Pretest

The tests and items were presented in a random order for each *S*. There were two items in each of the four conservation tests. Materials in the mass test were colored plasticine balls; in the liquid test, 2 identical beakers with equal amounts of water, a tall thin glass, and a short wide glass; in the number test, 2 sets of five black checkers; in the length test, 2 yellow sticks, 10 inches each in length.

At the start of each test, stimuli were presented so that they were perceptually and quantitatively equal. Thus the 2 sticks were aligned horizontally so that their ends matched, the checkers were placed in one–one correspondence, equal amounts of water were poured into two like beakers, and the round balls of plasticine were the same. Then *S*s were asked if the stimuli had the "same or different amount"; *S*s who judged the stimuli as "same" watched *E* rearrange or alter one of the stimuli and were asked a series of questions. An example of questions asked after each transformation is:

(1) "Do these have the same or different amounts of clay?"
(2) (If *S* answered "different.") Does this one (the altered object) have more or less clay?" and "Why do you think so?"
(3) (If *S* first answered "same.") "Why do you think so?"

Similar questions were used in all conservation tests, except the wording was changed as required by the particular test. Table 1 summarizes the items used in both pre- and posttests.

Nonconservation Classification

The *S*'s answers to the questions and his explanations were used jointly to define nonconservation. Explanations given to "same" responses were rated as either "adequate," "inadequate," or "ambiguous." A response was rated adequate if it referred to former equality, reversibility of the transformation, compensation, addition and subtraction, the irrelevancy of the transformation, and partition or matching schema. Examples of adequate responses observed are; "you just moved it (the stick), and if you moved it back they'd be the same length," "because it's water and will go down further and be wider in this jar, but stays the same amount," or "it's the same number because you didn't put another one there." A response was rated inadequate if it was "magical," provided no information, or made reference to perceptual cues or events in the experiment. Examples of these are "my mommy told me," "I don't know," "they are the same size," or "you made this into a pencil." Any response that could not be rated as adequate or inadequate was

designated ambiguous. All explanations given to a "same" answer were rated independently by two judges and the percent of agreement for 149 correct answers was 95.6, indicating high inter-rater reliability.

An *S* was defined as a nonconserver if he gave (1) no correct answer or explanations or (2) one or two correct answers, but no correct explanations for any of the right items. On this basis, 70 children were defined as nonconservers. Sixty-six met the first criterion, four the second. Ten of these were lost due to illness on training days. The *S*s were tested until a squad of 12 nonconservers was observed; 4 *S*s from each squad were randomly assigned to one experimental condition. This procedure continued until 20 *S*s were assigned to each condition.

TABLE 1
DESCRIPTION OF TEST ITEMS ON PRE AND POSTTEST OF CONSERVATION

Conservation test	Item	Pretest	Item	Posttest
Length	(1)	one stick moved on the horizontal to *S*'s right	(1) + (2)	same as pretest
	(2)	one stick placed at the center and vertical to standard stick	(3)	V's placed at the end of variable stick so as to induce Muller-Lyer illusion. The V's pointed inward and made the stick look shorter
			(4)	V's pointed outward and made the stick look longer
Number	(1)	one row spread out to look longer	(1) + (2)	same as pretest
	(2)	one row made into a circle	(3)	one row moved together to look shorter
			(4)	one row divided into two groups: One of three and one of two
Liquid	(1)	contents of one beaker poured into a tall thin glass	(1) + (2)	same as pretest
	(2)	contents of one beaker poured into a short wide glass	(3)	contents of one beaker poured into two smaller glasses
			(4)	contents of one beaker poured into a triangular shaped beaker
Mass	(1)	one ball made into a long thin sausage shape	(1) + (2)	same as pretest
	(2)	one ball made into a cross shape	(3)	one ball made into two smaller balls
			(4)	one ball made into a square

Oddity Control (OC) Training

The stimulus materials for oddity problems consisted of small, three-dimensional toys glued, face up, on 2 × 2-inch blocks of wood. Thirty-two sets of three blocks (two alike and one different) were made by combining 16 separate pairs of identical stimuli, so that each stimulus was included in four different sets; twice as the member of the identical pair and twice as the odd object. An example of a stimulus set would be two toy lions and one toy cup.

These 32 stimulus sets were used in 32 training problems, with 16 being presented per day for 2 consecutive days. Each problem consisted of six trials. The position (left, middle, or right) of the odd object within a problem was randomized with the restriction that the odd object occurred twice in each position. The *S*'s task was to point to either two objects that were the "same" or two that were "different." On half the trials, *E* asked *S* to point to two toys that were the "same" and on the other half to two that were "different." The latter response had to include a choice of the "odd" stimulus in order to be judged correct. Whenever *S* made a correct choice, he was told; "Yes, that is right, and here is a 'prize'." When wrong, he was told; "No, that is not right." A noncorrection procedure was followed. The prizes were trinkets.

Learning Set Training

Problems. There were 32 six-trial problems, 16 were length and 16 were number. Each problem consisted of three stimulus objects, two that contained identical, and one that contained different quantities (e.g., two rows of five chips versus one row of three chips, or two 6-inch sticks versus one 10-inch stick).

Thirty-two six-trial problems were used to assure that *S*s received extensive training with a large number of different examples of the relevant conservation principles. The choice of number and length concepts derived from examination of the nature of the problems. It has been noted (Piaget, 1952) that children often define numerosity in terms of length cues. For example, children say that the number of chips in a row increases when the row is made longer. Alternation between number and length problems here meant that sometimes the length was relevant and sometimes irrelevant. The interchange of number and length tasks was viewed as one way of forcing the child to see that a quantity cue can be either relevant or irrelevant, and that he has to discriminate when a particular cue is, in fact, relevant. To solve all problems, the child would have to learn to separate out the different cue functions of length, as well as, ignore irrelevant cues within a problem.[3]

[3] In regard to the above it should be noted that Beilin (1965) used a feedback

(a) *Between problem variations.* The changes that occurred between, but not within problems were color (red, green, yellow, or blue), size and shape (for length, large or small square or circular sticks; for number, large or small rectangular or circular chips), starting arrangements (horizontal, vertical, horizontal and vertical, and geometrically arranged rows of chips or sticks), and quantity combinations. For length,

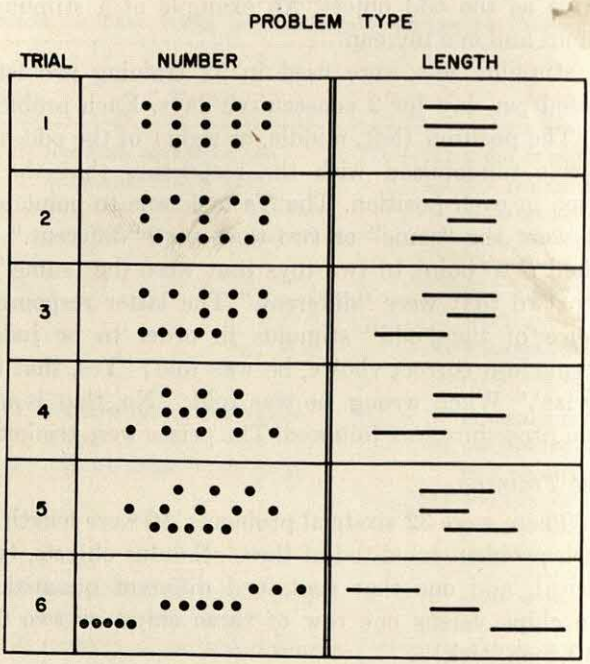


FIG. 1. Schematic representation of intraproblem variations for a length and number problem presented during SC and LS training.

the quantity combinations were: one or two of either 6- and 10- inch or 5½- and 7-inch sticks. The number quantity combinations were: one or two sets of four and six, or three and five objects. In both kinds of problems, variation in length and number values were used to assure generalized responding to length and number cues.

For each type of problem (length or number), 16 stimulus sets were constructed by combining these dimensions above. Over problems, each stimulus value occurred four times.

(b) *Within problem variations.* Examples of the stimulus variations

training procedure with length and number problems and failed to obtain transfer on conservation tests. This may be attributed to the fact that he used significantly fewer trials and presented the length and number problems in blocks rather than randomly.

within a problem are shown in Fig. 1. The stimulus sequences within each problem were designed to approximate certain features of the conservation tests and allow for later analyses of the kinds of errors made. The six trials were arranged as follows: On Trial 1, as at the beginning of conservation tests, the stimuli consisted of patterns where all cues were relevant and redundant. For example, two sticks of equal length were placed parallel to each other and with ends aligned, but the third different lengthed stick was not aligned nor necessarily parallel with the other sticks. Thus, a choice of the aligned and parallel sticks as "same" could be on the basis of either length (which is relevant) or end matches or parallel cues (which are irrelevant). Likewise, in the number problems, on Trial 1, the number, length, and parallel cues were redundant. Trials 2–5 served as "transformation" trials, where the alignment and geometric cues varied independently of length and/or number cues. On Trial 6, the stimuli were moved so as to hold constant irrelevant cues. For example, the 3 sticks were spatially separated and nonparallel. For *S* to respond correctly, he would have to do so on the basis of the relationship between quantities per se.

Although the questions asked and the stimuli shown in LS training were different than those in OC training, other features of training (e.g., feedback presentation and randomization) were identical to those in the OC condition. After the stimuli were arranged for a particular trial, *E* said "show me two sticks that are the same (or different) length" or "show me two rows that have the same (or different) number of things in them."

Stimulus Change (SC) *Control.* The stimuli and training procedures for the SC condition were identical to those for the LS condition, except no feedback was given. At the end of each session, *S* was told that he was playing the game "very well" but nothing else.

Transfer Tests. Each *S* was tested the day after training and then 2–3 weeks later on the four conservation tasks shown in Table 1. The administration procedure was identical to that used in pretesting, except for the addition of items in each test. When the conservation tests were readministered 2–3 weeks after the first posttest, the presentation order of tasks and items within each was rerandomized, but otherwise testing was the same.

RESULTS AND DISCUSSION

Oddity Training

The *S*s in Group OC made virtually no errors in training. The mean number of errors per *S* was 1.5 indicating that understanding of "same-different" was most likely acquired before training.

SC and LS Training

Figure 2 shows the proportion of correct responses for the 32 successive problems for the LS and SC groups. The first problem is shown by halves, since learning in the LS condition began within the first 3 trials; this was not true for the SC group. Inspection of Fig. 2 reveals that both groups had the same initial probability of a correct response (approximately .60). Learning Set *Ss* began to learn *immediately* and reached an asymptotic performance level of approximately 95% correct. An *S* was considered to have learned the length and number problems, if for each he had no more than one error for at least the last two problems. Nineteen of the 20 LS *Ss* reached this criterion for both types of problems.

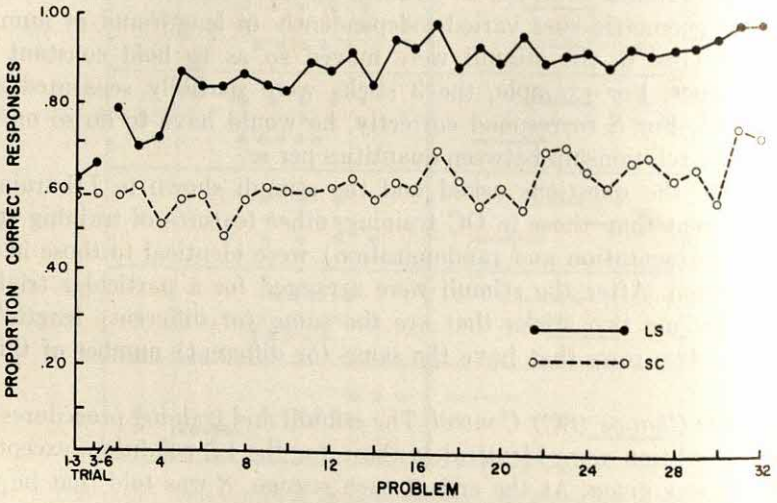


FIG. 2. Probability of a correct response per 6-trial problem for groups LS and SC.

In contrast, there was almost no learning about quantity in the SC condition. The average learning curve is relatively flat, rising over the last problems to .70 correct. Inspection of individual data revealed this improvement was contributed by 6/20 *Ss*. Four of these *Ss* "learned" both concepts, while the other two "learned" only the length concept. The remaining *Ss* failed to reach criterion or show any improvement over trials. Fig. 3 gives "learning" curves for both kinds of SC *Ss*, "learners" and "nonlearners." In comparing Figs. 2 and 3, note that *Ss* who came to respond to quantity in the SC condition did so much more slowly than LS *Ss*.

The LS acquisition data were analyzed separately for the length and number problems. The probability of a correct response on each problem

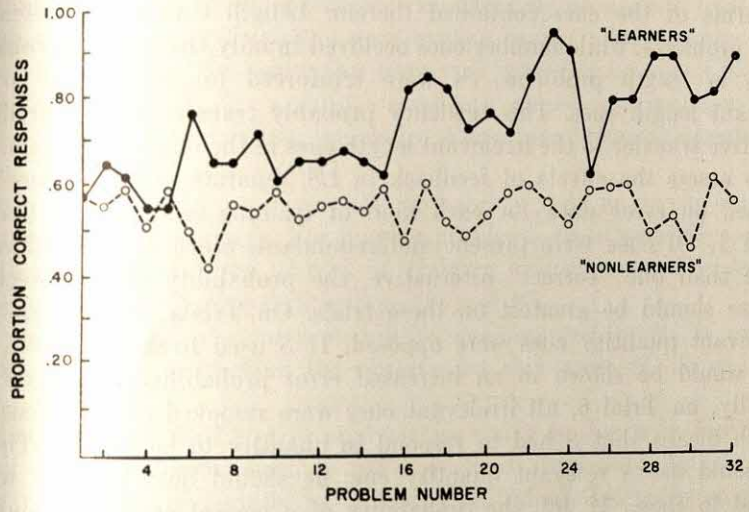


FIG. 3. Probability of a correct response per problem for "learners" and "nonlearner" in the SC group.

type during the course of training is given in Fig. 4. One can see that for problems 1–10 the learning curves do not differ in any systematic way. However, for problems 11–30, the probability of a correct response on length exceeds that on number in all but two cases. Thus, length concepts would appear to be "easier" to acquire than number ones. However, this result is understandable if one analyzes the two problems

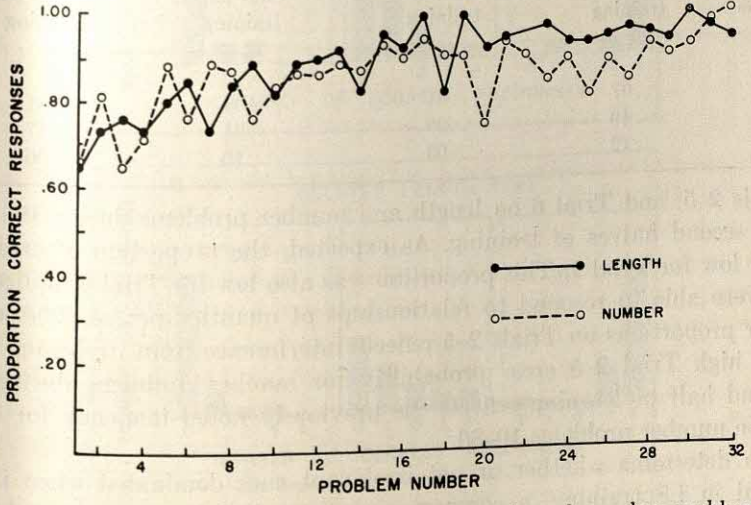


FIG. 4. Probability of a correct response on length and number problems for group LS.

in terms of the cues contained therein. Length cues were present in both problems, while number cues occurred in only the number problems. Thus in length problems, *Ss* were reinforced for responding to the relevant length cues. This tendency probably transferred and produced negative transfer to the irrelevant length cues in the number problem.

To assess the effects of feedback in LS, separate analyses were performed on error data for each kind of training trial. Recall that on Trial 1, all cues were present and redundant. Since Trial 1 offered *S* more than one "correct" alternative, the probability of a correct response should be greatest on these trials. On Trials 2–5 relevant and irrelevant quantity cues were opposed. If *S* used irrelevant cues, then this would be shown in an increased error probability on Trials 2–5. Finally, on Trial 6, all irrelevant cues were removed or held constant, which means that *S* had to respond to quantity to be correct. Thus if he could use a relevant quantity cue, he should quickly learn to respond to these. If not, the probability of a correct response should be close to chance ($p = .60$).

Table 2 gives the proportion of errors group LS made on Trial 1,

TABLE 2
ERROR PROBABILITIES DURING THE FIRST AND SECOND HALVES OF LS TRAINING

	Proportion of errors			
	Length		Number	
Problem type	First half training	Second half training	First half training	Second half training
Trial				
1	.07	.01	.05	.04
2–5	.19	.08	.21	.17
6	.12	.03	.10	.06

Trials 2–5, and Trial 6 on length and number problems during the first and second halves of training. As expected, the proportion of errors is very low for Trial 1. This proportion was also low for Trial 6, indicating *Ss* were able to respond to relationships of quantity per se. The larger error proportions on Trials 2–5 reflects interference from irrelevant cues. The high Trial 2–5 error probability for number problems during the second half of learning reflects the previously noted tendency for *Ss* to err on number problems 10–30.

To determine whether or not irrelevant cues dominated when introduced in LS training, a comparison was made between the number of errors due to use of such irrelevant cues versus errors due to random or position responses. In scoring the errors, it was assumed that *S*'s choice

of stimuli indicated to which cue he was attending on a given trial. For example, in Fig. 1, on Trial 3, if an *S* made a "same" choice of the rows containing three and five chips, his error was judged to be one of matching ends; on Trial 5, for a similar erroneous choice, spacing cues were judged to be those controlling the choice. For a more detailed description of this scoring, see Gelman (1967).

Given that an error occured, the probability that an irrelevant cue was used is extremely high. For number problems, these were .89 and .82 for the first and second halves of training, respectively. For length problems, these same probabilities were .88 and .85.

A similar pattern of error probabilities was observed for SC *S*s, except these were larger, reflecting the observation that most SC *S*s did not learn.

On the basis of the LS error analysis, we conclude that with feedback, young children quickly learn to use a quantity dimension. In fact, the rapid acquisition (especially as indicated on Trial 6) strongly suggests that *S*s had some preexisting understanding of quantitative relationships. Nevertheless, when irrelevant cues were introduced, they were frequently

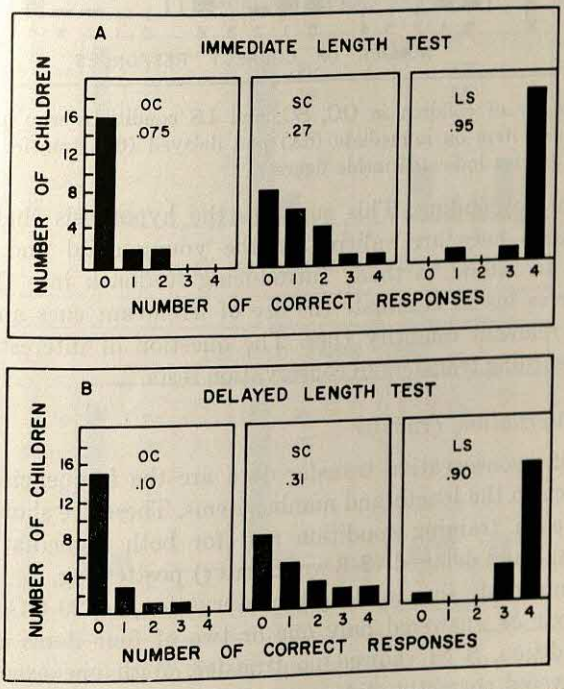


FIG. 5. Number of children in OC, SC, and LS conditions who correctly answered 0, 1, 2, 3, or 4 length items on immediate (5A) and delayed (5B) transfer tests. (Overall proportions correct indicated inside figures).

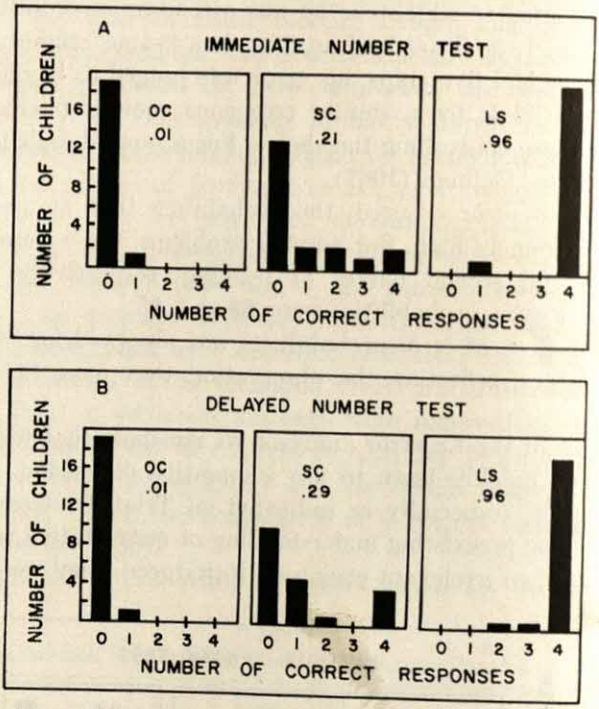


FIG. 6. Number of children in OC, SC, and LS conditions who answered 0, 1, 2, 3 or 4 number item on immediate (6A) and delayed (6B) transfer tests. (Overall proportion correct indicated inside figures).

the basis for responding. This supports the hypothesis that irrelevant nonquantitative cues are salient for the young child and that he is more likely to attend to them. Introducing feedback into the task apparently forces his to eliminate the use of irrelevant cues and to attend to and use relevant quantity cues. The question of interest is whether or not this learning transfers to conservation tests.

Specific Conservation Transfer

The specific conservation transfer data are the frequencies of correct answers given to the length and number items. These are shown in Figs. 5 and 6 for each training condition and for both immediate (one day after training) and delayed (2–3 weeks later) posttesting.

On the immediate test for length conservation, 16/20 OC *Ss* failed to improve. Four *Ss* answered only one or two of four items correctly. In the SC condition, 8 *Ss* showed no transfer. With one exception, those *Ss* who answered correctly, did so to only one or two items. In contrast, all but two LS *Ss* made perfect scores on length tests. Comparing Figs.

5 and 6, one can see that the pattern of results on the immediate number tests was virtually identical to that on length tests.

The overall percentages correct for specific length and number tests were 95 and 96, 27 and 21, 7.1 and 1.3 for the LS, SC, and OC conditions, respectively.

Specific tests administered two to three weeks later yielded essentially the same results as above, indicating durability of the training effects.

The SC data were analyzed in terms of those *Ss* who did and did not "learn" during training. The 6 SC "learners" contributed 60% of the correct answers. This is to be compared with the 95% correct responding by LS *Ss*, suggesting that the "learning" by SC *Ss* was not durable.

Nonspecific Transfer

The results for immediate and delayed tests on liquid are summarized in Fig. 7; those for mass are given in Fig. 8.

As on specific tests, *Ss* in group OC showed no transfer to nonspecific tests. More notably *Ss* in the SC condition also failed to transfer to non-

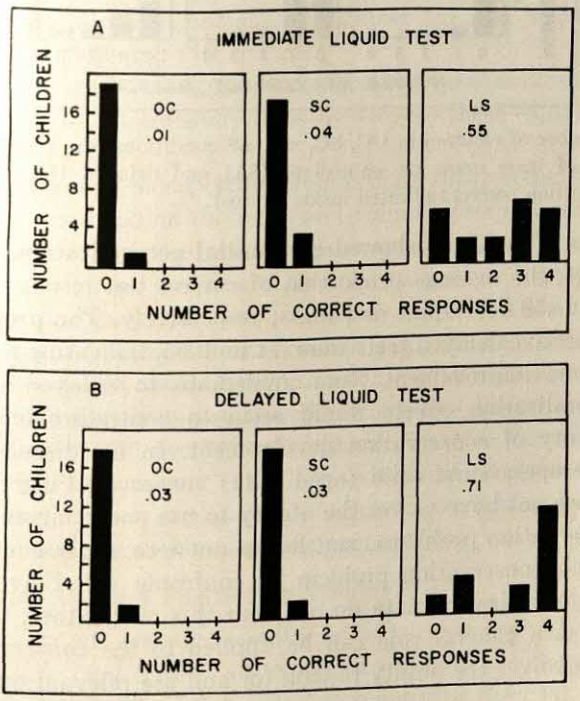


FIG. 7. Number of children in OC, SC, and LS conditions who correctly answered 0, 1, 2, 3, or 4 liquid items on immediate (7A) and delayed (7B) transfer tests. (Overall proportion correct indicated inside figures).

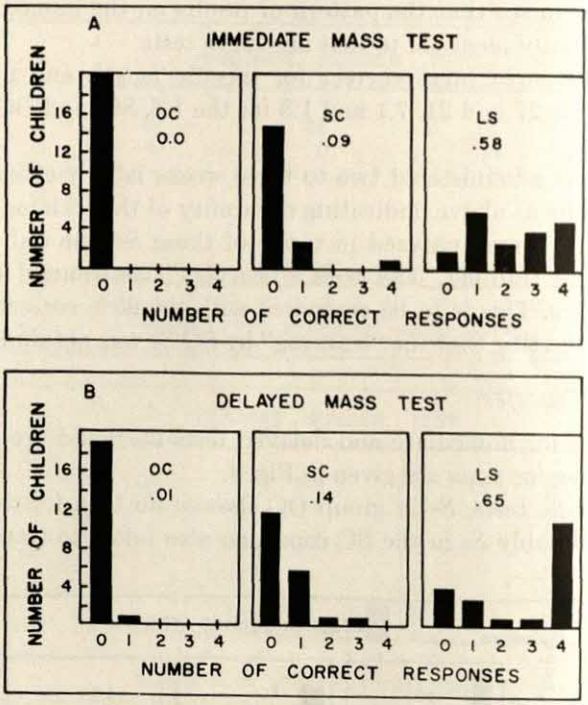


FIG. 8. Number of children in OC, SC, and LS conditions who correctly answered 0, 1, 2, 3, or 4 mass items on immediate (8A) and delayed (8B) transfer tests. (Overall proportion correct indicated inside figures).

specific tests. The LS *Ss* showed substantial generalization. On the immediate tests, the average proportion of correct test responses by LS *Ss* were .55 and .58 for liquid and mass, respectively. The proportions for the two respective delayed tests were .71 and .65, indicating retention and perhaps, some improvement from immediate to delayed tests. These strong generalization effects would seem to contradict one aspect of Piaget's theory of conservation development. In his discussion of how the concrete operational child (aged 7–11) conserves, Piaget states that the child does not have as yet the ability to use one common set of rules for all conservation problems, and learns anew to apply concrete operations to each conservation problem he confronts (cf. Flavell, 1963; p. 204). The generalization data do not bear this out; rather, they support the idea that a general rule can be applied to the conservation tasks, one which involves the ability to look for and use relevant quantity cues.

Explanations on Transfer Tests

Following the procedure used for pretest data, all explanations for correct responses were rated as adequate, inadequate, or ambiguous. The

percent agreement for two independent ratings of 602 explanations was 96.2.

Regardless of test type or test interval, LS *Ss* consistently gave more adequate explanations than their SC (or OC) counterparts. For LS *Ss*, the conditional probability of an adequate explanation, given a correct answer on the immediate test, was .97, .78, .78, and .91 for number, length, mass, and liquid, respectively. The respective probabilities on delayed tests were .95, .86, .89, and .97. Some examples of adequate explanations which illustrate the extent to which LS *Ss* conserved on posttests are; "you have to break them if you are going to change the length," "you haven't taken any away," "it's (water) lower but wider and so the same," "they were the same before and you can see if you make it (plasticine) back," or "it doesn't matter if you do that." This good correspondence between responses and explanations was not observed for the SC *Ss*. Approximately 27% of all specific transfer items were answered correctly by SC *Ss*, but of these, less than 50% were explained adequately. These *Ss* gave 39.4% adequate explanations on immediate and 47.9% on delayed tests. Comparing kinds of specific tests, SC *Ss* gave a higher percentage of adequate explanations for number (61 and 73 for immediate and delayed) than length (19 and 24).

Since there were so few explanations given to nonspecific transfer items by SC *Ss*, and to all transfer items by OC *Ss*, these explanations were not analyzed.

A consideration of nonspecific explanations sheds some light on why generalization occurred on the mass and volume tests following LS training. Aside from having learned to ignore specific context cues in length and number problems in LS training, *Ss* may have learned to ignore some nonspecific classes of cues (cf. Harlow, 1959; Restle, 1958). Such a general class might be stimulus change, regardless of the quantitative cue. Also LS *Ss* might have learned to maintain an initial judgment from Trial 1 through the stimulus changes. That this kind of learning did occur, is shown in nonspecific explanations like "you haven't done anything to change the amount" or "they used to be the same and so they have to be now." Such explanations occurred frequently; in fact, every LS *S* gave at least one such nonspecific explanation.

Immediate and Delayed Posttest Differences

Note that for all transfer conditions, save one, both the number of correct responses and quality of explanations improved from the first to the second posttest. The LS *Ss* showed some forgetting on length tests. To evaluate the reliability of these changes, t tests for correlated observations were performed on all measures. Nonsignificant t scores were obtained for all but one test; the increase in number of "same"

responses given by LS *Ss* on liquid tests is significant, but only when a 1-tail test is used ($t = 1.857$; $p < .05$). Thus it would seem that the observed changes are not reliable.

Overview

Starting with an analysis of conservation as a problem in attention and discrimination, the present research has shown that, given appropriate training, one can elicit conservation behavior from children who initially fail to conserve on classical conservation tests. Appropriate training seems to involve two factors: (1) An opportunity to interact with many different instances of quantitative equalities and differences and (2) feedback, which presumably tells *S* what is and what is not relevant to the definition of quantity. This is supported by training and transfer results from both SC and LS *Ss*. The SC *Ss* received only changing stimuli, while the LS *Ss* received both changing stimulus experience and feedback. Some of the SC *Ss* learned to conserve in a limited way. There was a small amount (27%) of specific generalization, but almost no *nonspecific* transfer. In contrast, with LS training almost perfect *specific* and considerable *nonspecific* transfer occurred. In addition, LS *Ss* were better able to explain their correct answers. Finally, it seems that LS training brought *Ss* to use a general rule like "it doesn't matter what you do or pay attention to the way it is to start."

There are two alternative interpretations that might be made of these results. The first involves a statement about the nature of concept acquisition per se. We might say that the LS procedure approximated the actual conditions under which a child learns to conserve *de novo*. However, the data go against this interpretation in several ways. First, LS *Ss* very quickly mastered the training task. If they were learning to define quantity and quantitative invariance *de novo*, we might expect more errors than observed and certainly would not expect learning within the first six trials. Since there was very fast acquisition, it seems more appropriate to say the five year old can work with quantity if "told" to do so. Feedback seems to be a very effective way of communicating the task requirements to a young child.

A second result of interest corroborates this interpretation. Analyses of Trial 6, which provided stimuli containing only the relevant quantity cues, showed that *Ss* had little or no difficulty in responding to quantity. Even during the first half of training *Ss* were able to work directly with the relevant cue. In fact, the only trials which provided a source of difficulty were those where irrelevant perceptual cues were present. These error analyses support the hypothesis that a young child can respond correctly to quantitative relationships, but has to be "instructed" on their

use (cf. Braine and Shanks, 1965). The five-year-old child apparently does have to learn to respond *consistently* to quantity and not be distracted by irrelevant cues, but does not have to learn, *de novo,* to define quantity and invariance. That *S*s adopted somewhat general rules may reflect new learning. Still, other investigators have indicated that young children know more about quantitative relationships than demonstrated in conservation tests, e.g., it has been shown that *S*s can accurately anticipate quantitative changes (Taponier, 1962; cited by Berlyne, 1965), or predict reversible effects (Berlyne, 1965; Wallach *et al.,* 1967), and still not conserve. It could be that these responses are present in a child's repertoire, but are dominated by strategies under the control of irrelevant stimuli. If so, training which extinguishes the use of irrelevant cues should also bring out the "correct" verbal responses.

It should be noted that our basic results lend support to the positions of Wallach *et al.* (1967) and Zimiles (1963). Wallach *et al.,* have suggested that conservation depends on a child learning to ignore misleading cues; and Zimiles (1963) offers an explanation of number conservation in terms of *S* learning to ignore spatial cues, while developing a number "set." Also of interest is a recent experiment by Kingsley and Hall (1967) who have successfully trained length and weight conservation with LS procedures. Although these findings agree with those of the present study, it is not clear from an analysis of the training procedures used by Kingsley and Hall why transfer was obtained. Kingsley and Hall's use of LS training differs from that of the present study. Rather than specifically training *S*s to ignore irrelevant and attend to relevant aspects of the conservation problem, they trained *S*s on a graded series of subtasks related to conservation (e.g., appropriate use of scales and then the operation of addition and subtraction). It is possible that the experience on these subtasks involved ignoring the irrelevant features therein, and hence, indirectly served the same purpose as the LS experience in the present study.

A critical question often raised about the present kind of research is whether or not the LS *S*s "really" have conserved on post-tests. Insofar as our tests measure conservation, the answer is yes. That *S*s were able to generalize and give logical explanations on nonspecific tests seems most significant. However, Piaget (1967) might argue that such tests are not sufficient and that our LS *S*s are "pseudoconservers" and would fail to perform correctly on other acts such as pouring an equal amount of water into large, narrow and shallow, wide jars. This can only be answered with further studies, since the present design did not include such tests.

Aside from the above, the results point to two lines of research. One

calls for the development of techniques that will yield specific descriptions of what the five-year-old knows about quantitative relations, and how his behavior differs from older children. For example, the data support our assumption that young children operate on a hierarchy of quantitative response strategies and define quantity multidimensionally. What seems called for is a careful analysis of this hierarchy and how it changes. A second line of investigation points to the need to develop techniques that could be used with younger *Ss*. The present research does not tell us how the child is able to respond to relevant quantity cues or what such a statement means operationally. Put differently, if by the time the child is five he has formulated the hierarchies we postulate, we still have the tasks of tracing and explaining their development.

REFERENCES

ALMY, M., CHITTENDEN, E., AND MILLER, P. *Young children's thinking: Studies of some aspects of Piaget's theory*. Teacher's College Press, Columbia Univ., New York, 1966.

BEILIN, H. Learning and operational convergence in logical thought development. *Journal of Experimental Child Psychology*, 1965, **2**, 317–339.

BERLYNE, D. E. *Structure and direction in thinking*. Wiley, New York, 1965.

BRAINE, M. D. S. Piaget on reasoning: A methodological critique and alternative proposal. In W. Kessen and C. Kuhlman (Eds.), Thought in the young child. *Monographs of the Society for Research in Child Development*, 1962, **27**, No. 2.

BRAINE, M. D. S., AND SHANKS, B. L. Conservation of a shape property. *Canadian Journal of Psychology*, 1965, **19**, 197–207.

BRUNER, J. S., OLVER, R. O., AND GREENFIELD, P. M., *et al. Studies in cognitive growth*. Wiley, New York, 1966.

FLAVELL, J. H. *The developmental psychology of Jean Piaget*. Van Nostrand, Princeton, New Jersey, 1963.

GELMAN, R. S. *Conservation, attention and discrimination*. Unpublished doctoral dissertation, University of California, Los Angeles, 1967.

HARLOW, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science*. Study 1, Vol. 2. *General systematic formulations, learning and special processes*. McGraw-Hill, New York, 1959.

INHELDER, B., AND PIAGET, J. *The growth of logical thinking from childhood to adolescence*. Basic Books, New York, 1958.

KINGSLEY, R. C., AND HALL, V. C. Training conservation through the use of learning sets. *Child Development*, 1967, **38**, 1111–1126.

LAWRENCE, D. H. The nature of a stimulus: Some relationships between learning and perception. In S. Koch (Ed.), *Psychology: A study of a science*. Study II, Vol. 5. *Process areas*. McGraw-Hill, New York, 1963.

PIAGET, J. *The child's conception of number*. W. W. Norton, New York, 1952.

PIAGET, J. A review of studies in cognitive growth by J. S. Bruner, R. R. Olver, P. M. Greenfield *et al. Contemporary Psychology*, 1967, **12**, 532–533.

RESTLE, F. Toward a quantitative descriptlon of learning set data. *Psychological Review*, 1958, **65**, 182–191.

TRABASSO, T., AND BOWER, G. *Attention in Learning: Theory and research*. New York: Wiley and Sons, Inc., 1968.

WALLACH, L., WALL, A. J., AND ANDERSON, L. Number conservation: The roles of reversibility, addition-subtraction, and misleading perceptual cues. *Child Development*, 1967, **38**, 425–442.

WALLACH, M. A. Research in children's thinking. In H. W. Stevenson (Ed.), Child psychology: Sixty-second yearbook, *National Society for Studies in Education*. Part 1. Univ. of Chicago Press, Chicago, Illinois, 1963.

ZIMILES, H. A note on Piaget's concept of conservation. *Child Development*, 1963, **34**, 691–695.

ZIMILES, H. The development of conservation and differentiation of number. *Monographs of the Society for Research in Child Development*, 1966, **31**, No. 6.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 188–194 (1969)

Setting Similarity and Successive Discrimination Learning by Children[1]

BONNIE J. LUBKER
Institute of Child Behavior and Development, University of Iowa

The experiment was designed to test the theoretical prediction that a decrease in the similarity of the settings, achieved by increasing the number of nonspatial dimensions varying between settings, decreases the difficulty of the successive discrimination problem. Three successive problems, differing as to the number of nonspatial dimensions that varied between settings, were administered to children. The pairs of stimuli in the two settings of Problem 3 differed with respect to three dimensions. They differed on two dimensions in Problem 2. One dimension varied between settings in Problem 1, the standard successive discrimination problem. As predicted, performance was highest on Problem 3, intermediate on Problem 2, and lowest on Problem 1. However, only the difference between Problems 3 and 1 was statistically significant.

The stimuli to be discriminated in the standard successive discrimination problem typically differ with respect to a nonspatial dimension and the spatial dimension. Both values on the spatial dimension and one of the two values on the nonspatial dimension are present in a setting. The two settings differ with respect to the value on the nonspatial dimension present in each setting. Both values on the nonspatial and spatial dimensions are rewarded and nonrewarded equally often. Stimulus compounds, each consisting of a value on the nonspatial dimension and on the spatial dimension, are differentially rewarded. The two settings are presented in a prearranged random order, and *S* is required to select the stimulus that is consistently rewarded in each type of setting.

A comparison of the two settings in the standard successive discrimination problem indicates they are similar in that the pairs of stimuli differ with respect to only one nonspatial dimension. A decrease in the similarity between the settings in the successive problem can be achieved by increasing the number of nonspatial dimensions on which the pairs of stimuli differ. An extension of Hull-Spence theory that specifies a rule for combining the generalized and conditioned habits and inhibitory

[1]The cooperation of Ralph Delozier, principal of Mark Twain School, and Walter Schnelle, principal of Henry Longfellow School, and their staffs is gratefully acknowledged. Gratitude also is extended to Pamela Parris for collecting the data.

tendencies of stimulus components in multidimensional stimuli, the Hypothesis of Stimulus Interaction (Spiker, 1963), predicts that a decrease in the similarity of the settings decreases the difficulty of the successive problem. The present experiment was designed to test this prediction. Three successive discrimination problems, differing as to the number of nonspatial dimensions that varied between settings, were administered to different children. The pairs of stimuli in the two settings of Problem 3 differed with respect to three nonspatial dimensions (i.e., form, size and brightness). They differed on two nonspatial dimensions (e.g., form and brightness) in Problem 2. One nonspatial dimension (e.g., form) varied between settings in Problem 1, the standard successive discrimination problem. According to the Stimulus Interaction Hypothesis, performance should be highest on Problem 3, intermediate on Problem 2, and poorest on Problem 1.

METHOD

Problems. The arrangements of stimuli used in each of the three successive discrimination problems are shown in Fig. 1. The stimuli differed in form, brightness, size,[2] and position.

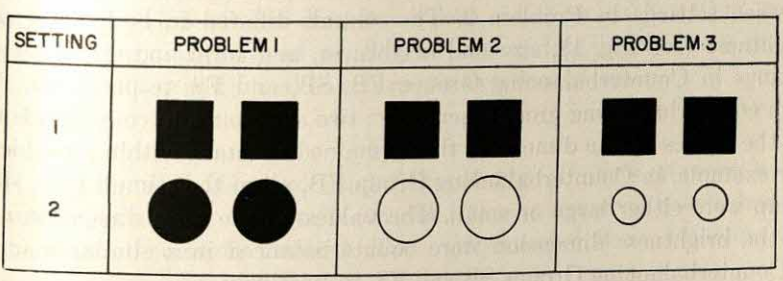


FIG. 1. The stimuli presented in the three successive discrimination problems. Selection of the stimulus on the left was rewarded in Setting 1 in each problem; selection of the right stimulus was rewarded in Setting 2.

In Problem 1, the standard successive discrimination problem, the two forms (square and circle), and the two positions (left and right) were scheduled for reward and for nonreward equally often. The stimu-

[2] Manipulation of the size dimension in the successive problem is to be avoided since a difference in the size of the stimuli between the two settings results in a corresponding difference in the distinctiveness of the spatial stimuli between settings. The solution of the successive problem requires the discrimination between the spatial stimuli as well as between the nonspatial and, according to the Stimulus Interaction Hypothesis (Spiker, 1963), an increase in the distinctiveness of the spatial stimuli, achieved through a greater spatial separation of the nonspatial stimuli, would result in an improvement in performance on the successive problem (see Spiker and Lubker, 1965).

lus compounds, left square and right circle, were consistently rewarded. Three groups were constituted to counterbalance for the nonspatial stimulus dimension that varied between settings in Problem 1. The stimuli differed in form in Counterbalancing Group F (see Fig. 1). In Counterbalancing Groups B and S, the stimuli differed in brightness and size between settings, respectively (not shown in Fig. 1). There were four subgroups to counterbalance for the values on the two nonspatial dimensions that were constant within a problem in each of these three main counterbalancing groups. All of the stimuli in a subgroup were either large and black, large and white, small and black, or small and white in Counterbalancing Group F. In Counterbalancing Group B the two values on the form and size dimensions were counterbalanced in a similar manner. The values on the form and brightness dimensions were similarly counterbalanced in Counterbalancing Group S.

In Problem 2 the two forms (square and circle), brightnesses (black and white), and positions (left and right) were rewarded and nonrewarded equally often. The black square on the left and the white circle on the right were consistently rewarded. There were three groups to counterbalance for the two nonspatial stimulus dimensions that varied between settings in Problem 2. The stimuli differed in both form and brightness (see Fig. 1), size and brightness, and form and size between settings in Counterbalancing Groups FB, SB, and FS, respectively. For each counterbalancing group, there were two subgroups to counterbalance for the values on the dimension that remained constant within a problem. For example, in Counterbalancing Group FB, all of the stimuli for a subgroup were either large or small. The values on the form dimension and on the brightness dimension were counterbalanced in a similar manner in Counterbalancing Groups SB and FS, respectively.

All *Ss* assigned to Problem 3 were given the problem shown in Fig. 1. Three nonspatial dimensions varied between settings in this problem. The two forms (square and circle), brightnesses (black and white), sizes (large and small), and positions (left and right) were equally often rewarded and nonrewarded. The large black square on the left and the small white circle on the right were consistently rewarded.

Apparatus. The stimuli were wooden blocks. Each block was halved along the frontal plane with the two halves joined by a hinge. A marble placed in the interior of a block was revealed by raising the top half of the block. There were four blocks each of the following sizes, brightnesses, and forms: large black square, large white square, small black square, small white square, large black circle, large white circle, small black circle, and small white circle. The dimensions of the blocks were as follows: 3.5 inches (large square); 3 inches (small square); 4-inch diameter (large circle); and 3-inch diameter (small circle).

Two blocks, the centers of which were 10 inches apart, were placed on an 18 × 22-inch red turntable on each trial. A 13 × 22-inch vertical screen in the center of the turntable prevented *S* from observing *E* as she prepared the apparatus for the next trial.

Subjects. The *S*s were 108 second-grade children from the Iowa City public schools, ranging in age from 6 years, 11 months to 9 years, 1 month. Thirty-six *S*s were assigned randomly to each of the three successive discrimination problems.

Procedure. The *S*s were tested individually. The following instructions were given: "This is a test to see if you can *learn* to find a hidden marble. When you learn the secret, you will be able to find the marble on each turn. [*S* is shown a red triangular block.] This is the way to look for a marble. [*E* presents the blocks appearing on the first trial.] Go ahead and look in the block that you think has the marble. Open only one each time." When *S* found a marble for the first time he was instructed to put it in the can placed on his right.

Each *S* received 120 noncorrection trials. A trial consisted of the presentation of one of the two settings. The two settings in a problem were presented equally often in a prearranged random order within each block of ten trials. Eight stimulus blocks, four for each setting, were used during the training of an *S*. The appropriate blocks were paired at random from trial to trial to control for extraneous cues.

RESULTS

The means and standard deviations of the numbers of correct responses obtained in 120 trials on the three successive discrimination problems are shown in Table 1. The rank order of these problems was

TABLE 1
MEAN NUMBERS OF CORRECT RESPONSES AND STANDARD DEVIATIONS FOR THE SUCCESSIVE DISCRIMINATION PROBLEMS

Problem	Mean CR	*SD*
1	63.03	11.86
2	69.81	19.84
3	75.81	20.56

consistent with the rank order predicted by the Stimulus Interaction Hypothesis (Spiker, 1963). A test (Jonckheere, 1954) of the predicted rank order against the null hypothesis resulted in a rejection of the latter beyond the .04 level of confidence.

Figure 2 shows the mean numbers of correct responses obtained on the three problems for each of the twelve blocks of ten trials. A statistical analysis (Type I design; Lindquist, 1953, pp. 267 ff.) was per-

formed on the correct-response data. The between-subjects factor was problems and the within-subjects factor was trial blocks.

The analysis indicated that the main effect of problems was statistically significant ($p < .025$). In order to determine which problems differed significantly from each other, t tests were conducted using an error term based on the variability of all three problems. There was a significant difference between the means for Problems 1 and 3 ($p < .01$) but the differences between Problems 1 and 2 and Problems 2 and 3 were not statistically significant.

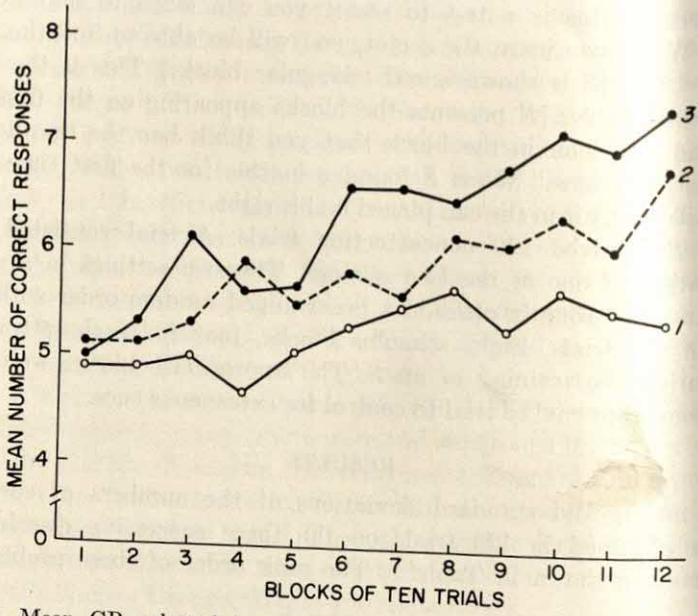


FIG. 2. Mean CRs plotted in blocks of ten trials obtained for the successive discrimination problems.

There was a nonsignificant interaction between problems and trial blocks but the main effect of trial blocks was significant ($p < .001$), indicating that the mean number of correct responses for the three problems combined increased with training. Further analyses were conducted to determine whether there was a statistically significant increase in the mean number of correct responses between trials 1–10 and trials 111–120 for each problem. Related t tests of the difference between the first and last trial blocks for Problems 2 and 3 were significant ($p < .001$). The relatively small increment in performance between the first and last trial blocks in Problem 1 resulted in a nonsignificant t statistic, indicating that the *Ss*, given the standard successive discrimination problem, performed near the chance level throughout training.

An analysis (Simple-randomized design; Lindquist, 1953, pp. 47 ff.) of the correct-response data obtained for Counterbalancing Groups F, B, and S in Problem 1 indicated that the counterbalancing effect was not significant. A similar analysis for Counterbalancing Groups FB, SB, and FS in Problem 2 revealed a significant counterbalancing effect ($p < .05$). Performance was higher in Group SB than in Groups FB and FS with the result that t tests of the difference between Groups SB and FS ($p < .02$) and Groups SB and FB ($p < .05$) were significant whereas the difference between Groups FS and FB was not.

DISCUSSION

The results were consistent with the theoretical prediction that a decrease in the similarity of the settings, achieved by increasing the number of nonspatial dimensions varying between settings, decreases the difficulty of the successive discrimination problem. Performance was highest on Problem 3, intermediate on Problem 2, and lowest on Problem 1, the standard successive problem. However, only the difference between Problems 1 and 3 was statistically significant. Similar results were obtained by Croll (Exp. II, 1967) in an experiment involving three-position oddity discrimination problems that differed as to the number of nonspatial dimensions on which the rewarded stimulus was odd. In Problem 3 the odd stimulus in each setting differed from the other two stimuli on three nonspatial dimensions. The odd stimulus in each setting in Problem 2 was odd with respect to two dimensions. In Problem 1, the standard oddity problem, the odd stimulus in each setting was odd with respect to only one dimension. Croll found that an increase in the number of nonspatial dimensions, on which the odd stimulus was odd, decreased the difficulty of the oddity problem, although only the difference between Problems 3 and 1 was statistically significant.

The finding that the *S*s given the standard successive problem performed near the chance level throughout training was surprising in view of the relatively large number of trials administered. The remarkable difficulty for children of the standard successive problem, when the settings are presented in a random order and *S* is required to respond directly to the stimulus source, has been demonstrated in experiments by Lipsitt (Exp. II, 1961), Spiker and Lubker (1965), and Lubker (Exp. I, 1967). The results of these experiments were inconsistent with respect to the solvability of the successive problem by children, which is probably due to differences in the age of the *S*s and the number of training trials given. The successive problem was so difficult for the kindergarten and first-grade children in the Lubker experiment that they failed to show appreciable learning in 72 trials. Even the fourth-grade children

in Lipsitt's experiment did not depart from chance performance throughout training on the successive problem; however, only 36 trials were given. Spiker and Lubker, on the other hand, observed consistent improvement in performance on the successive problem, even for those third- and fourth-grade children who failed to reach a criterion of seven out of eight correct responses within 56 trials.

REFERENCES

Croll, W. L. *Oddity discrimination learning as a function of the number of dimensions along which the correct stimulus is odd.* (Doctoral dissertation, University of Iowa) Ann Arbor, Mich.: University Microfilms, 1967. No. 68-915.

Jonckheere, A. R. A distribution-free k-sample test against ordered alternatives. *Biometrika,* 1954, **41,** 135–145.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Lipsitt, L. P. Simultaneous and successive discrimination learning in children. *Child Development,* 1961, **32,** 337–347.

Lubker, B. J. Methods of training successive and conditional discriminations. *Child Development,* 1967, **38,** 597–606.

Spiker, C. C. The hypothesis of stimulus interaction and an explanation of stimulus compounding. In L. P. Lipsitt and C. C. Spiker (Eds.), *Advances in child development and behavior.* Vol. 1. New York: Academic Press, 1963. Pp. 233–264.

Spiker, C. C., and Lubker, B. J. The relative difficulty for children of the successive and simultaneous discrimination problems. *Child Development,* 1965, **36,** 1091–1101.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 195–202 (1969)

Effects of Redundancy on Information-Reduction Tasks[1]

IOANNIS PARASKEVOPOULOS
University of Illinois

Thirty elementary school children sorted dot patterns that were either asymmetrical or symmetrical about (a) a vertical axis, (b) a horizontal axis, or (c) both. Asymmetrical patterns were sorted faster and with fewer errors than the symmetrical patterns ($p < .001$). Stimulus redundancy interfered with rapid visual discrimination presumably by reducing stimulus uniqueness and distinctness. When the complexity of the input was sufficient to force the *S*s to reduce the stimulus to distinctive parts, the retarding effects of redundancy on discrimination were in direct relation to the amount of redundancy and independent of the *S*s' chronological age.

Posner (1965) proposed a taxonomy of psychological tasks based upon the relation between input and output information required for perfect performance on the task. This taxonomy consists of three types of tasks: (a) Information-conservation tasks in which the *S* is required to preserve all of the input information in his response and any increase or decrease in information during transmission represents error; (b) information-addition tasks which involve information creation, so that the output information must exceed the input if the *S* is to perform the task; and (c) information-reduction tasks which require the *S* to produce a subset of the stimulus input. The loss of information does not represent error, but rather is necessary to produce the required output.

Paraskevopoulos (1967, in press) found that, in information-conservation tasks, performance is a function of the amount and the form of stimulus redundancy as well as the *S*s' chronological age and intelligence. The present study investigated the effects of stimulus redundancy on information reduction tasks.

METHOD

Stimuli. The stimuli consisted of eight-dot patterns embedded in matrices. The procedure described by Attneave (1955) was adopted to

[1]The report is based upon a dissertation submitted in partial fulfillment of the requirements for doctoral degree at the University of Illinois. The author is indebted to the members of his dissertation committee, Professors H. W. Hake, S. Jones, S. A. Kirk, K. Scott, and M. Tatsuoka, for their valuable guidance in planning and conducting the study.

generate the patterns. The dots were arranged in the following four variations: (a) *Asymmetry*. The position of each dot was independent of the position of any other dot in the pattern (Fig. 1a); (b) *Bilateral Symmetry*. Only the dots in the left half portion of the pattern were determined independently. The right half portion of the pattern was a

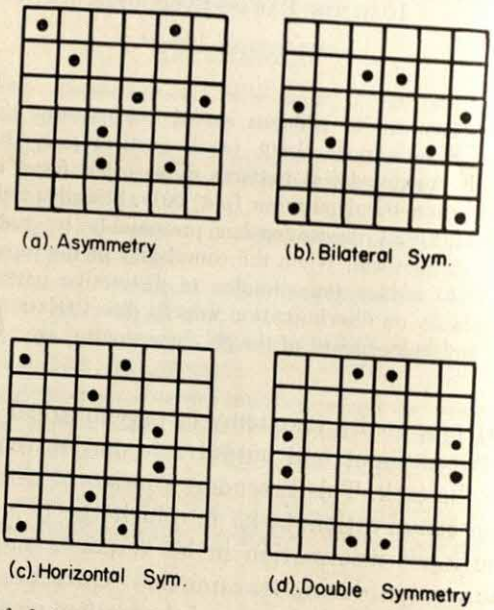


FIG. 1. Sample of dot patterns for asymmetry, bilateral, horizontal, and double symmetry.

mirror image of the left half (Fig. 1b); (c) *Horizontal Symmetry*. The bilateral patterns were turned 90°. The direction of turn was randomly determined. Thus, the bilateral and horizontal patterns were identical in all respects but in orientation (compare Fig. 1b with 1c); and (d) *Double Symmetry*. Only the positions of the dots in the upper left quadrant were determined independently. The rest of the dots were placed so that they were symmetrical both horizontally and vertically (Fig. 1d). For each mode, seven different patterns were generated and six black and white photographic copies for each pattern were obtained. One copy was mounted on cardboard to be used as a model in the sorting task. The other five copies were pasted on playing cards; these comprised the deck to be sorted.

Apparatus. Seven rectangular boxes, one for each model card, were put together side by side. Above these boxes a panel was installed to hold the cardboard with the model cards. A photograph of the apparatus, as viewed by the *S*, is shown in Fig. 2.

Procedure. Each *S* was tested individually. Sitting in front of the apparatus, the *S* was presented with the model panel and corresponding deck of cards. To counterbalance carryovers of fatigue, practice or boredom effects, each *S* was presented with a different permutation of the four modes.

The examiner instructed the *S* to sort the cards by matching them with the model cards as quickly as possible. Following any questions by the *S*, the instructions for speed and accuracy were reemphasized. For each *S*, the time for sorting and the errors in matching were recorded for each mode.

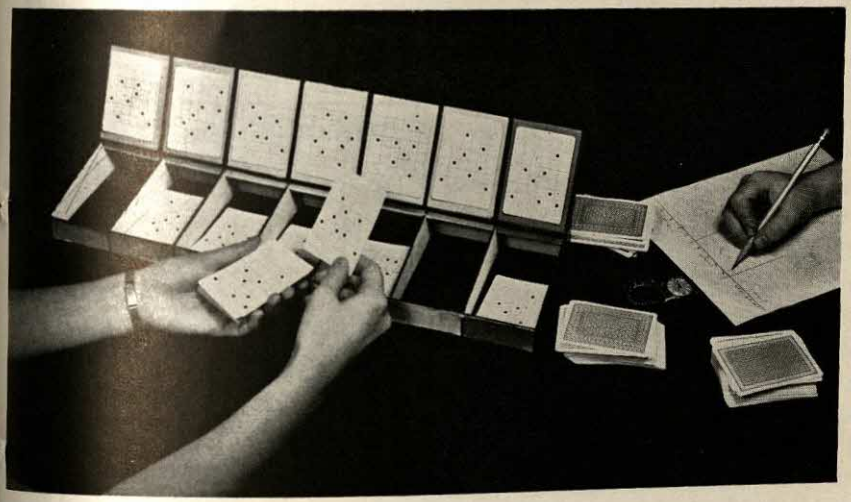


FIG. 2. Testing apparatus.

Statistical analysis. Time and errors were analyzed by multivariate analysis of variance with repeated measures. To stabilize the variance the time scores were replaced by their reciprocals and the error scores by their square roots (Edwards, 1963). The nature of the differences was explored by means of discriminant analysis (Jones, 1960).

EXPERIMENT I

Subjects. Twenty-four children attending upper elementary grades served as *S*s.

Results. The mean time and error scores are presented in Fig. 3. The variance and covariance matrices for the transformed scores and the mean squares and cross-products matrices are given in Table 1 and 2. The differences between the centroids of the four modes was statistically significant ($F_{6/182} = 330.29$; $p < .001$). To explore the source of this difference, discriminant analysis was performed.

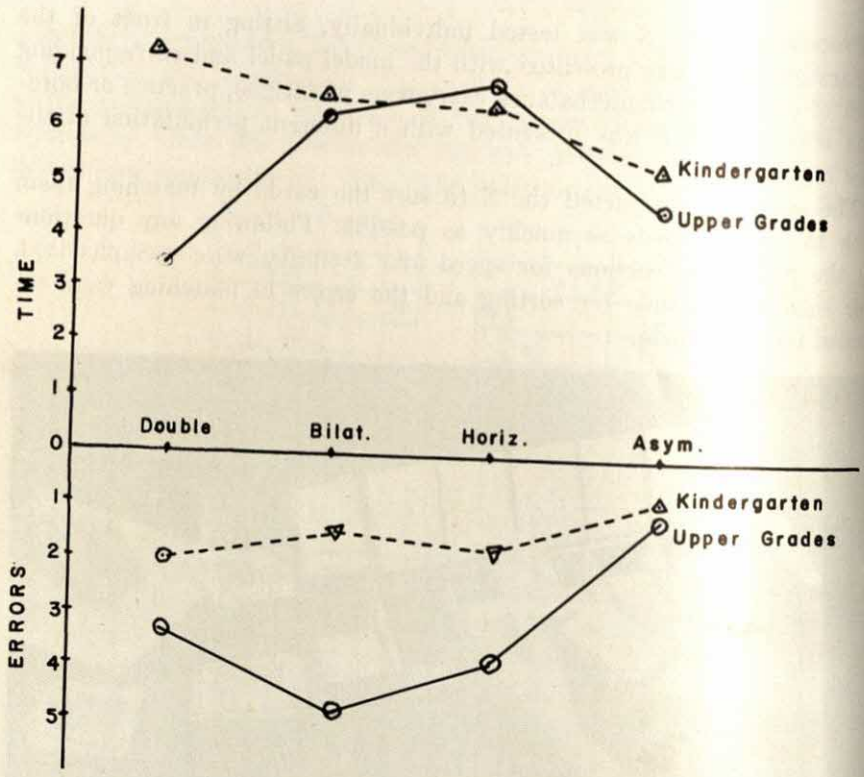


FIG. 3. Mean time and error scores of the four modes for upper elementary grades and kindergarten children.

TABLE 1

VARIANCE AND COVARIANCE MATRICES OF TIME AND ERROR SCORES[a]

Mode	Upper elementary grades		Kindergarten	
	Time	Errors	Time	Errors
Double	.3437		.1062	
	−.3053	88.9711	−.9554	20.6654
Bilateral	.1999		.2364	
	.1612	70.2260	−1.0911	12.8869
Horizontal	.2128		.0877	
	−.0499	79.8115	−1.5824	60.9876
Asymmetrical	.3071		.1529	
	1.0597	28.6244	−.7525	9.7684

[a] All numbers have been multiplied by 100.

TABLE 2

MEAN SQUARES AND CROSS-PRODUCTS MATRICES OF TIME AND ERROR SCORES[a]

Variables	Upper elementary grades			Kindergarten		
	Subjects			Subjects		
Time	.6213		*ndf*	.6360		*ndf*
Errors	1.4620	170.1207	23	−5.3420	68.1553	5
	Modes			Modes		
Time	2.9016		*ndf*	.2929		*ndf*
Errors	−33.7038	595.8721	3	−2.0710	16.8222	3
	Error term			Error term		
Time	.1628		*ndf*	.0213		*ndf*
Errors	.3185	36.3828	69	.0281	19.0049	15

[a] Mean squares and cross-products values have been multiplied by 100.

The first discriminant function accounted for 93% of the variance. Each variable accounted for equal portions of the discriminatory power of the function; time accounted for 51% and errors for 49%. This discriminant function was $Vu = .0642\ (\text{Errors}) - .09979\ (\text{Time})$.

Large scores on a discriminant function indicate that performance is relatively high on variables with positive coefficients and relatively low on variables with the negative coefficients. Inversely, small discriminant scores indicate that performance is relatively low on variables with positive coefficients and high on variables with negative coefficients. The time scores, however, were transformed by reciprocal transformation; thus, small transformed scores signify large scores in the original scale. Therefore, large scores on the obtained discriminant function indicate both large Time and Error scores; inversely, small discriminant scores indicate both small Time and Errors scores.

Figure 4 presents the order of the mean discriminant scores for the

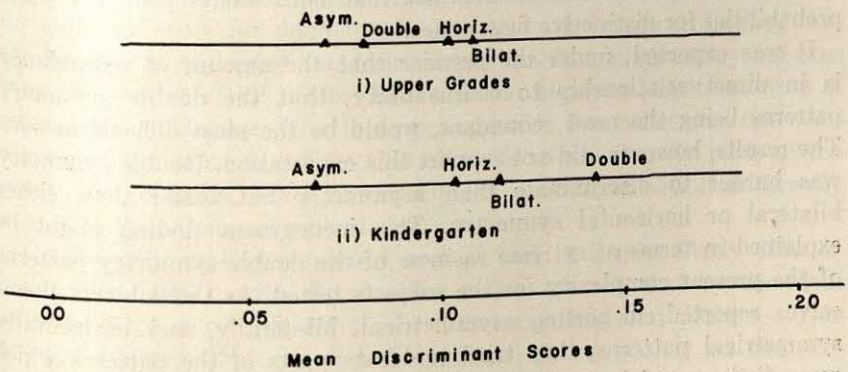


FIG. 4. Mean discriminant scores of the four modes for upper elementary grades and kindergarten children.

four modes. The largest mean discriminant score was for bilateral and horizontal symmetry; the smallest was for asymmetry. These results suggest that the time spent, and the errors made sorting the asymmetrical patterns were the smallest of the four modes. The longest time spent and the most errors made were in sorting the bilateral and horizontal patterns.

Discussion. The results suggest that, unlike with memory tasks (Paraskevopoulos, in press), symmetry retards performance on discrimination. Similar retarding effects of the stimulus redundancy were observed in experimental situations involving reaction time with visual stimuli (Gregg, 1954) and recognition of visual forms and word lists (Anderson and Leonard, 1958; Dale and Baddeley, 1962; Deese, 1956).

Discrimination probably requires the selection of but few unique characteristics of each stimulus (information-reduction task) which reliably differentiate one pattern from the others. This sampling strategy has been observed in several experimental situations (Forsman, 1966; Munsinger, 1965). The more irregular a stimulus is, the more unique characteristics it presents for selection. Randomness provides more and easier discriminable combinations of details and, thus, permits the *S* to reduce the amount of information he must process before he can respond correctly. Symmetry, with the constraints it introduces, limits the degrees of freedom for pattern variations, thus increasing the similarity and subsequent confusability. Therefore, the retarding effects of redundancy on discrimination can be attributed to the differences in homogeneity between symmetrical and asymmetrical patterns. The nondifferential performance on bilateral and horizontal symmetry bears out the notion of within-mode pattern homogeneity. The bilateral and horizontal patterns in the present experiment were identical in all respects but in orientation. The practically equal performance on these two modes might be due to the fact that both modes provided equal probabilities for distinctive figure cues.

It was expected, under the premise that the amount of redundancy is in direct relationship to confusability, that the double symmetry patterns being the most redundant, would be the most difficult to sort. The results, however, did not support this expectation. Double symmetry was harder to discriminate than asymmetry but easier than either bilateral or horizontal symmetry. This incongruent finding might be explained in terms of extreme easiness of the double symmetry patterns of the present complexity for the subjects tested. As the subjects themselves reported, in sorting asymmetrical, bilaterally, and horizontally symmetrical patterns, they tried to locate parts of the patterns which were distinct and to match the patterns in terms of these parts. But, in classifying the double symmetry, they "worked" with the whole pattern.

It seems that the present complexity was not sufficient to force the subjects to encode only parts while soring double symmetry. The information load of the double symmetry with 8 dots was far below the subjects' capacity. In another experiment Paraskevopoulos (1967) found that the mean errors in recall for these patterns was negligible. To test the hypothesis of light information load two alternatives were offered: (a) To construct more complex double symmetry patterns (probably 16-dot patterns) and to administer them to the same subjects; or (b) to use the stimuli of the present study with younger children. The second alternative was followed and carried out in Experiment II.

EXPERIMENT II

Subjects. Six 5-year-old children served as *S*s. In a recall experiment (Paraskevopoulos, in press) it was found that kindergarten children made substantial number of errors in reproducing from memory 6-dot patterns and that the mean error differences among the four modes were statistically different. It was, therefore, assumed that the load of the 8-dot patterns was far beyond the capacity of kindergarten children.

Results. The mean error and time scores are presented in Fig. 3. The variance and covariance matrices and the mean squares and cross-products matrices are presented in Tables 1 and 2. The differences between the centroids of the four modes were statistically significant ($F_{6/38} = 20.18$; $p < .001$).

Discriminant analysis yielded the function $Vk = .0094$ (Errors) + 1.000 (Time) accounting for 99% of the variance. The mean discriminant scores for the four modes are presented on Fig. 3. The smallest mean discriminant score was, as in the case of the upper elementary school children, for asymmetrical patterns indicating that the least time and fewest errors were made in sorting asymmetrical patterns. The mean discriminant score for double symmetry was the largest indicating that the time spent and errors made in sorting the double symmetry patterns were the largest of all modes. Bilateral and horizontal patterns were of intermediate difficulty. The findings suggested that whenever the complexity of the input is sufficient to allow *S*s to decode only parts of the input, the retarding effect of redundancy on discrimination is directly related to the amount of redundancy.

Consistently in Experiments I and II and in a pilot study (Paraskevopoulos, 1967) with 8-year-olds it was found that symmetrical patterns were more difficult to sort than asymmetrical patterns. This findings suggests that the retarding effects of redundancy on the performance of information reduction tasks is independent of the *S*s' chronological age.

REFERENCES

ANDERSON, N. S., AND LEONARD, J. A. The recognition, naming, and reconstruction of visual figures as a function of contour redundancy. *Journal of Experimental Psychology*, 1958, **56**, 262–270.

ATTNEAVE, F. Symmetry, information, and memory for patterns. *American Journal of Psychology*, 1955, **68**, 209–222.

DALE, H. C. A., AND BADDELEY, A. D. Alternatives in testing recognition memory. *Nature Lond.*, 1962, **196**, 93–97.

DEESE, J. Complexity of contour in the recognition of visual form. USAF WADC TR 56-60, February, 1956.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Holt, 1963.

FORSMAN, R. G. Age differences in the effects of stimulus complexity and redundancy on pattern discrimination in a visual search task. Unpublished doctoral dissertation, University of Illinois, 1966.

GREGG, L. W. The effect of stimulus complexity on discriminative responses. *Journal of Experimental Psychology*, 1954, **48**, 289-297.

JONES, L. V. Some illustrations of psychological experiments designed for multivariate statistical analysis. The Psychometric Laboratory, University of North Carolina, Chapel Hill, N. C., 1960.

MUNSINGER, H. Tachistoscopic recognition of stimulus variability. *Journal of Experimental Child Psychology*, 1965, **2**, 186–191.

PARASKEVOPOULOS, I. N. Symmetry: its effects on recall, preference and discrimination of visual pattern in relation to age and intelligence. Unpublished doctoral dissertation. University of Illinois, 1967.

PARASKEVOPOULOS, I. N. Symmetry, recall and preference in relation to chronological age. *Journal of Experimental Child Psychology*, 1968, **6**, 258–264.

POSNER, M. I. Memory and thought in human intellectual performance. *British Journal of Psychology*, 1965, **56**, 197–215.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 203–219 (1969)

Modification of the Classroom Behavior of a Disadvantaged Kindergarten Boy by Social Reinforcement and Isolation[1]

SALLY A. SIBLEY,[2] MARTHA S. ABBOTT, AND BETTY P. COOPER
Duke University

Reinforcement techniques were employed to decrease the disruptive, resistant, and assaultive behaviors and to increase the appropriate social interaction of a white, economically disadvantaged, kindergarten boy. The treatment program involved presentation of teacher attention contingent upon "desirable" classroom behavior, withholding of attention contingent upon "inappropriate" behavior, and social isolation contingent upon "unacceptable" behavior.

The subject's behavior was classified according to the *Coping Analysis Schedule for Educational Settings* (CASES) (Spaulding, 1967), and the teacher's interactions with the subject were categorized according to their content. The subject's behavior and the teacher's interactions with him were recorded simultaneously on an event recorder by an observer who had attained high reliability.

The baseline data supported the teacher's contention that the subject was a behavioral problem, particularly in a strictly structured situation. The program was carried out daily in the activities of free-play, discussion, and rest. Some progress was made under the original program, but the subject's inappropriate behaviors were further decreased when they were punished (with isolation) rather than ignored. Extinction of the treatment program (i.e., decreased positive and neutral interactions contingent upon desirable behavior and withdrawal of isolation as an ultimate contingency) was introduced to demonstrate that the teacher's interaction was indeed a major controlling variable. After the successful extinction, the treatment was reinstated with favorable results. Although this description is accurate of the over-all treatment and its effects in general, there were differences in the program and in its effects in the various activities.

Time checks were made several weeks after the termination of the official study that indicated that the teacher was maintaining the treatment as an integral part of the child's environment and that he was still responding favorably.

[1] Supported by the Education Improvement Program (Ford Foundation), Durham, North Carolina.

[2] Research Training Fellow with the Duke University Center for the Study of Aging and Human Development (Grant 5T01HD00164 from the National Institute of Health and Human Development).

In recent years, interest has increased in the application of operant conditioning theory and techniques to the behavior problems of children. Reinforcement theory and the area of behavior therapy or modification focus on understanding a child's behavior as a function of the present environmental consequences of his behavior and the child's past interactions with his environment. Ullman and Krasner (1965) presented a series of case studies in this area, the majority of which focused upon deviant behaviors in children and schizophrenic adults. Staats (1964) has compiled a number of studies that extend conditioning principles to complex human behavior, mainly verbal behavior, communication, and social learning.

Most demonstrations of operant techniques with human behavior have been in laboratory situations. However, several studies that illustrate the application of reinforcement principles to nursery and kindergarten children in classroom situations have been conducted. Harris *et al.* (1964) carried out a study using positive social reinforcement to substitute well-developed walking behavior for the regressed crawling of a 3-year-old girl. Teachers have used positive social reinforcement (adult attention) to help a child showing persistent and marked isolate behavior to achieve and maintain more play relationships with peers (Allen *et al.*, 1964). Another investigation was undertaken by Johnston *et al.*, (1966) employing a planned schedule of positive social reinforcement to promote vigorous physical activity of an inactive, uncoordinated 3-year-old boy. Brison (1966) used the technique of social extinction of nonverbal communication to encourage a kindergarten child to talk. Finally, Staats (1964) reported on a study comparing token reinforcement, social reinforcement, and no reinforcement of reading behavior of 4-year-old children. The results showed that when reading was reinforced, attentional and work habits were strong and new words were learned rapidly, whereas both types of behavior deteriorated when reinforcement was not forthcoming. Thus, this research indicated that reinforcement principles provide effective and desirable means for increasing desirable classroom behavior of children.

The present investigation was undertaken to determine whether a program of positive social reinforcement of desirable behavior and punishment could discourage the disruptive and resistant behavior of a 5-year-old economically deprived boy. The kindergarten teacher used systematic presentation and withholding of her attention, as well as selective social isolation, to encourage a decrease in the child's aggressive, negative attention-getting, and resisting behaviors and an increase in his social, cooperative, and conforming behaviors. A secondary goal was to explore the applicability of a comprehensive behavior classification system in a

behavior-modification study. A third goal was to explore the possibilities of closely observing and recording teacher behavior in interaction with the subject and of employing these observations as a feedback to the teacher of the effect of her actions.

METHOD

Subject

"Bobby" was one of 12 children from low socioeconomic families enrolled in a kindergarten class. "Bobby," a white child in a class composed of an equal number of white and Negro children, was 5½ years of age at the start of the study.

When "Bobby" entered kindergarten, he was described by his teachers as a bright, alert, verbally skilled, and physically well-coordinated child. After refusing several times to enter the testing situation, "Bobby" did cooperate with the psychometrist during the experimental phase of this study. On the Peabody Picture Vocabulary Test he achieved an IQ of 105, on the Columbia Mental Maturities Test, 104, and on the Stanford-Binet, 103, placing him well within the average range.

The teachers requested a special study using behavior modification techniques after more traditional techniques had proved ineffective with "Bobby's" disruptive and resistant behavior. Quite frequently he would move away from the kindergarten group and proceed to disrupt the ongoing activity. To a disturbing degree, "Bobby" was physically and verbally assaultive toward the other children. On several occasions, "Bobby" had sudden and "uncontrollable" outbursts. He resisted his teacher's attempts to calm him, whined and cried, "Leave me alone," and threatened, "My daddy will come beat you up." The social and family history was obtained and led to the picture of a young boy protected and undisciplined by his mother and physically overwhelmed by his father.

Behavioral Categorization and Recording

The scale employed to analyze "Bobby's" behavioral change was the *Coping Analysis Schedule for Educational Settings* (CASES) developed by Spaulding (1967). The scale consists of 13 basic categories, which were further classified for the purpose of this study into desirable, inappropriate, and unacceptable behavior, as presented in Table 1.

One observer was responsible for the recording of the subject and teacher behavior throughout the study. This observer compared her classifications with that of four other trained observers during 12 ten-minute time-sampling checks conducted across the different stages of the experiment. The percentages of exact agreement (i.e., each CASES category)

the observer obtained with the four other observers ranged from 70% to 92%. The percentage of larger category agreement (i.e., desirable, inappropriate, unacceptable) ranged from 86% to 100%. There is, therefore, ample evidence that the system of behavioral classification employed in this study does have meaning and reliability, after training, beyond the lone observer.

TABLE 1

MODIFIED COPING ANALYSIS SCHEDULE FOR EDUCATIONAL SETTINGS (CASES)[a]

DESIRABLE (D)

5a *Self-Directed Activity*—working independently on an activity or project with interest

6 *Paying Rapt Attention*—listening and attending with interest to the ongoing activity

7a *Sharing and Helping*—contributing ideas and interests, volunteering answers, and helping others

8a *Social Interaction*—mutual interaction through conversation, games, and joint projects

9 *Seeking Support, Assistance, and Information*—asking for help, sympathy, and attention from teacher or peers

10 *Following Instructions Passively*—conforming to expectations without great interest

INAPPROPRIATE (I)

5b *Self-Directed Activity*—5a., but at an inappropriate time

7b *Sharing and Helping*—7a., but at an inappropriate time

8b *Social Interaction*—8a., but at an inappropriate time

11 *Observing Passively*—being distracted from ongoing activity

12 *Responding to Internal Stimuli*—no observable interaction with environment

UNACCEPTABLE (U)

1 *Assaultive Behavior*—direct verbal or physical attacks or destruction of property

2 *Negative (Inappropriate) Attention-Seeking Behavior*—loud or annoying disruptive behavior which seems to be directed toward obtaining the attention of others through unacceptable behavior

3 *Manipulating and Directing Others*—bossing others

4 *Resisting Authority*—actively or passively refusing to comply with teacher's expectations or requests

13 *Flight*—leaving the authorized limits of travel

[a] Spaulding, Robert L. *An introduction to the use of the coping analysis schedule for educational settings (CASES).* © 1967.

As presented in Table 2, the teacher's interactions with the pupil were classified according to their explicit content. As with the categorization of the subject's behavior, reliability checks of the teacher interaction were carried out with an independent observer. No attempt was made to calculate agreement on the onset and cessation of an interaction. Rather, the two observers agreed that an interaction was ongoing and independ-

ently rated the character of the interaction as positive, negative, neutral, or re-directing. Two comparisons were made, each consisting of 40 interactions, between this experiment's observer and one other observer. The first reliability check yielded an exact agreement percentage of 92.11% and the second, obtained during a different experimental condition, yielded an exact agreement percentage of 92.50%. Although the sample of reliability computations is small, the high agreement does indicate that this descriptive division of teacher/pupil interaction has meaning beyond the single observer.

TABLE 2

CLASSIFICATION OF TEACHER INTERACTIONS

Neutral: Conversation or relevant proximity without a connotation of explicit approval or disapproval
Positive: Verbal or nonverbal communication with explicit approval
Negative: Verbal or nonverbal communication with explicit disapproval or displeasure

All data were recorded on an event recorder and analyzed to show total time in each category, total frequency in each category, and percentage of time in each category during the activity. The total time and percentages were also computed for the more molar classifications of desirable, inappropriate, and unacceptable behavior. The teacher interactions were charted according to the type and length of interaction and "Bobby's" behavioral change during the interaction.

Procedure

Baseline. "Bobby's" behavior was recorded using the *Coping Analysis Schedule for Educational Settings* in order to ascertain the baseline character of his behavior. The reinforcing teacher's interactions were recorded simultaneously with "Bobby's" behavior. At first, data were taken during all of the kindergarten activities and then three activities were chosen for observation and intervention. Freeplay (30 minutes), discussion (10 minutes), and rest (5 minutes) were the activities selected due to the extent and variability of "Bobby's" undesirable behavior during these activities and the constancy of length of time from day to day.

After observation of "Bobby's" behavior in interaction with the teacher during the baseline period, several hypotheses were made. The teacher was positively reinforcing, and thus maintaining "Bobby's" negative and aggressive behavior, by her attention, although disapproving. This disapproval was too mild to serve as punishment, and no powerful punishing consequences followed the teacher's threats, so that teacher disapproval had not acquired conditioned punishment properties. The

subject was receiving peer social reinforcement for his aggressive and disruptive behavior. It was also hypothesized that isolation from peer and teacher attention would be punishing to "Bobby."

Social reinforcement—Treatment I. The social (potential) reinforcement was presented on a variable-ratio schedule, the exact values of which were not calculated, to give "Bobby" maximum possible adult attention contingent upon desirable behavior and minimum attention contingent upon inappropriate or unacceptable behavior, as defined in Table 1. The reinforcement schedule was carried out systematically by only one of the three teachers in the classroom, and an effort was made to hold all other variables constant throughout the study. Thus, the reinforcing teacher gave positive social attention to "Bobby" for desirable behavior, ignored all inappropriate behavior, and ignored unacceptable behavior unless it was intolerable, at which time the subject was given a short, negative, verbal threat of isolation. If he did not stop his unacceptable behavior within 5 to 10 seconds after a warning, "Bobby" was put in isolation for 5 minutes.

The isolation condition meant that he sat by himself in an enclosed cubicle in a room adjoining the kindergarten. "Bobby" was initially warned that if he were unable to sit there quietly by himself (the teacher returned to the classroom immediately), he would go to the principal's office to sit for 10 minutes. If "Bobby" continued his unacceptable behavior, the principal was to inform the teacher and then "Bobby" would be taken home. He was not informed of this final ultimatum because it was hypothesized that he might prefer to go home and, therefore, misbehave. It never was necessary to take him home. The purpose of the teacher's warning and isolation procedure was to develop teacher verbal disapproval as conditioned punishment.

During the study, the contingencies of social reinforcement were slightly altered, producing Treatment I and Treatment II. The first treatment consisted of ignoring all inappropriate behavior and ignoring all unacceptable behavior unless it was intolerable. "Bobby" emitted very little unacceptable behavior, but slightly increased his inappropriate behavior under this program. It was hypothesized that variables other than teacher attention were maintaining the inappropriate behavior that led to the initiation of Treatment II.

Social reinforcement—Treatment II. The reinforcing teacher continued to give positive social attention to "Bobby" for desirable behavior, but ignored inappropriate behavior only until it became disruptive to the group. Then he was given a verbal warning, followed by isolation if he did not behave within the desired limit. The second phase of treatment involved, therefore, less stringent requirements for punishment so that the teacher was, in effect, less tolerant.

Extinction. This stage was a brief period in which the reinforcing teacher attempted to return to baseline conditions, as nearly as possible, by interacting with "Bobby" with disapproval when he was exhibiting inappropriate and unacceptable behavior, but giving no verbal warnings and no periods of isolation. She decreased her positive and neutral interactions with the subject.

In general, the return to the baseline pattern of interactions was extinction or withdrawal of the reinforcement and punishment contingencies of the treatment program. This procedure was considered necessary to ascertain whether the teacher's attention and isolation from attention were the significant independent variables. The hypothesis was that if the teacher were the controlling variable, "Bobby's" behavior would become more undesirable and inappropriate during the extinction condition as compared to the treatment conditions.

Reintroduction of social reinforcement. During this stage, the teacher returned to the reinforcement schedule of Treatment II. The reinforcement schedule was gradually shifted from continuous to more intermittent until "Bobby" finally received adult attention in an amount normal for the group. After the completion of the study, data were taken on two days to check on the maintenance of the behavioral modification. An informal attempt was made to generalize the treatment to the two other teachers who were interacting with "Bobby."

Changes from one stage to the next condition were instituted according to a preset criterion of two consecutive days in which "Bobby's" percentage of desirable, inappropriate, and unacceptable behavior fell within the range of the preceding percentages of that behavior in each of the three activities. Because instability of behavior was one of "Bobby's" prime characteristics prior to the study, a criterion of stable behavior on consecutive days was held to be untenable during the baseline. The preset stability criterion was not enforced during the extinction condition because the trend was obviously accelerated in the expected direction and the resumed disturbance to the class motivated the teachers to strongly advocate resumption of the treatment.

RESULTS

Interaction Analysis

In a study of this kind, the dependent variable is the child's behavior and the independent variable is the teacher's behavior. In most investigations, the teacher's (or the therapist's) behavior is only described rather than tracked as closely and objectively as the subject's. In the present study, records of the teacher's behavior in interaction with the subject were maintained and employed as a feedback mechanism to inform the teacher of her success in carrying out the prescribed procedures.

The matrices in Table 3 represent the analysis of the teacher's interactions with the subject. All three activities have been combined. As indicated in the final column of Table 3, the teacher averaged nine interactions daily with the subject during Baseline. The column sums show that during Baseline, the teacher made an almost equal mean number of responses to the subject's inappropriate and unacceptable behavior combined as she did to his desirable behavior. The row sums indicate that the teacher emitted more negative interactions than she did neutral and

TABLE 3
MEAN DAILY FREQUENCY OF INTERACTIONS AND INTERACTION CONTINGENCIES BY EXPERIMENTAL CONDITIONS

		Desirable	Inappropriate	Unacceptable	Total	Total $\bar{X}$
BASELINE	Neutral	2			2	
	Positive	2			2	
	Negative	1	1	3	5	
	Total	5	1	3		9
TREATMENT	Neutral	15	1	1	17	
	Positive	5			5	
	Negative		1	1	2	
	Total	20	2	2		24
EXTINCTION	Neutral	3		1	4	
	Positive	1			1	
	Negative			2	2	
	Total	4		3		7
REINTRODUCTION	Neutral	18		1	19	
	Positive	5			5	
	Negative		1	2	3	
	Total	23	1	3		27

positive ones combined. The interior matrix illustrates that negative interactions, were primarily contingent upon unacceptable behavior, whereas neutral and positive interactions were entirely contingent upon desirable behavior, as would be expected. The mean of one negative interaction contingent upon desirable behavior typically involved incidents in which the teacher responded to an undesirable behavior which had ceased by the time she mentioned it so that the subject had changed to a desirable behavior.

During Treatments I and II, the teacher increased her total mean interactions to 24. She responded to desirable behavior more than to

inappropriate and unacceptable behavior by a factor of 5, as compared to the equality of Baseline. Neutral and positive interactions were more frequent than negative by a factor of 11, whereas there were more of the latter in Baseline. The contingencies within the matrix are as expected, except that some neutral interactions were emitted contingent upon undesirable behaviors. However, the negative interactions contingent upon desirable behavior were eliminated.

Extinction, or withdrawal of treatment contingencies, consisted of an almost exact return to Baseline with regard to responses to desirable behavior as compared to inappropriate and unacceptable behavior. On the other hand, the teacher did not return to the high proportion of negative interactions. Extinction was in essence a decrease in neutral and positive interactions and the withdrawal of isolation as a consequence.

Reintroduction of the treatment program after extinction closely matched the interactions during Treatments I and II. The total mean frequency of interactions was higher but the proportions were similar. There were higher frequencies of interaction with desirable and unacceptable behavior than in Treatments I and II.

In general, the matrices show that mean frequency of teacher/pupil interactions was greater in Treatment and Reintroduction as compared to Baseline and Extinction. These differences consisted primarily of increases in neutral and positive interactions contingent on desirable behavior.

Overview of Modification

The analysis of the subject's behavior throughout the study indicates a definite increase of desirable and decrease of inappropriate and unacceptable behavior as a result of the experimental treatment. Hypothetically, if the teacher's behavior in interaction with the subject were a major controlling variable, a specific trend in "Bobby's" behavior should be evident across the five experimental conditions. The desirable behavior of the subject should increase from Baseline through Treatments I and II, decrease toward the baseline level during Extinction, and increase again during Reintroduction.

An overview of the data is presented in Fig. 1. Days in which isolation occurred are omitted from this analysis. During Baseline, the subject emitted a greater percentage of desirable behavior in freeplay than in discussion and rest, activities with more limiting restrictions and expectations. During Treatment I, there were increases in all three activities with the greatest gain manifested during rest. During Treatment II, involving punishment of inappropriate as well as unacceptable behavior, there were further increases in desirable behavior in all activities. Ex-

tinction had differential effects within the three activities in that there were more noticeable decreases in desirable behavior during discussion and rest than in freeplay. Although Extinction did not result in a return to the baseline level, the trend is definite and in the expected direction. During Reintroduction, both discussion and rest show the expected increase in desirable behavior. More detailed consideration of the data in each activity follows.

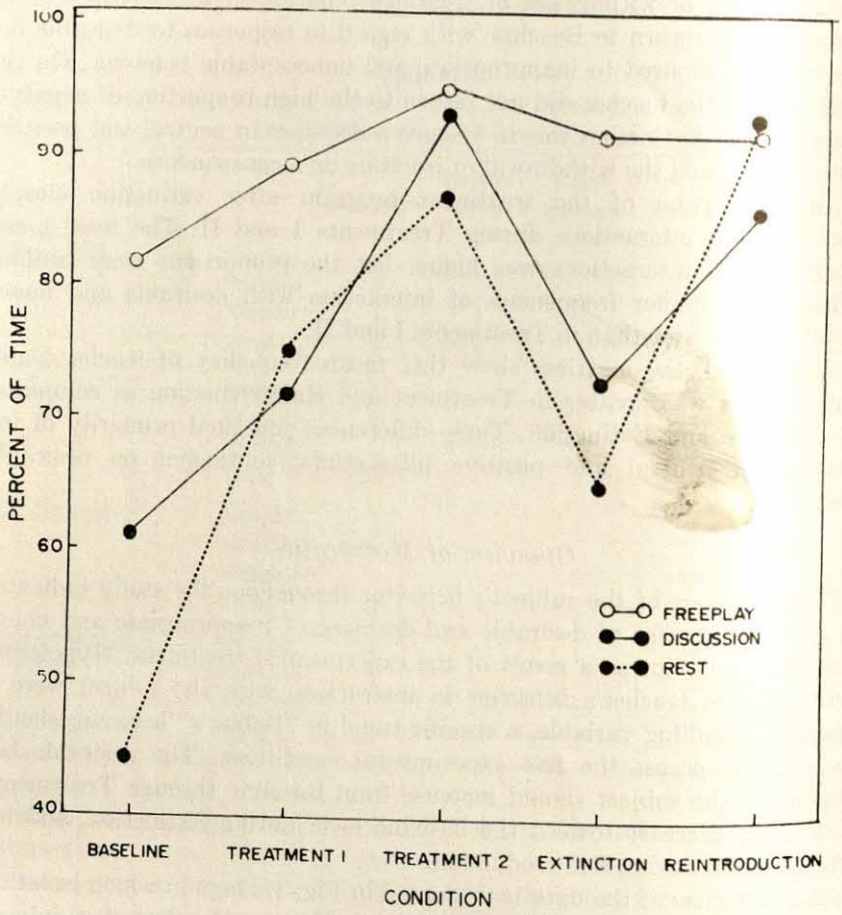


FIG. 1. Percentage of desirable behavior by activity and experimental condition.

Normative data from classes in the Education Improvement Program (EIP) on CASES are available in great detail. As compared to all boys in his own class, "Bobby's" desirable behavior was closest to the norm in all activities during Treatment II and Reintroduction. To compare with Fig. 1, the normative desirable behavior percentage for the boys in the class was 98% for freeplay, 92% for discussion, and 91% for rest.

Freeplay

"Bobby's" behavior was not a major problem during freeplay, even during Baseline, as compared to the other activities. Desirable behavior did increase as a function of the experimental treatment, although not as strikingly as in the other activities. There was no isolation during freeplay until Reintroduction; however, this fact does not exclude isolation as a factor since the same teacher was isolating the subject at other times. During Extinction, the subject did not completely revert to Baseline behavior, although there is some indication that his behavior was changing in the expected directions. During Reintroduction in freeplay, one of the more uncontrollable exceptions in the data occurred. The subject was involved in an unfortunate incident with a Negro adult that resulted in his self-imposed absence from school for one day. After this incident, racial name-calling (a major problem with him from the start) accounts for the sustained unacceptable behavior during Reintroduction.

In freeplay activity, it is interesting to note the change in "Bobby's" social behavior within the desirable category. During Baseline, 54% of his desirable behavior was self-directed activity (Category 5) and 42% was social interaction (Categories 7 and 8). It was hoped that desirable social behavior could be encouraged without sacrificing "Bobby's" self-directed behavior. In the social reinforcement conditions, the subject's self-directed behavior decreased to 46%, whereas social interaction increased to 50%. In Extinction, both of these desirable behaviors are about 47%. During Reintroduction, "Bobby's" social interaction again became the larger percentage with self-directed activity 43% and social interaction 55%, as in social reinforcement conditions.

"Bobby's" behavioral change within the unacceptable category during freeplay is also interesting to analyze. In looking at his total unacceptable behavior in Baseline, 51% was aggressive behavior (Category 1) and 25% was "bossy" manipulation of others (Category 3). During the two treatments, 73% of his unacceptable behavior was "bossy" manipulation and only 4% was aggressive. During Extinction, "Bobby's" aggressive behavior was again higher (38%) than "bossy" manipulation (21%), thus paralleling Baseline. This fact is especially interesting in that it provides evidence that, although total desirable behavior did not strikingly decrease during Extinction, more extreme unacceptable behavior did increase. In Reintroduction, both aggressive (45%) and bossy-manipulative (38%) behaviors were comparatively high, which was perhaps related to the aggressive incident outside of school.

Approximately 2 weeks after the final day of Reintroduction, data were taken during all three class activities to provide a partial check on

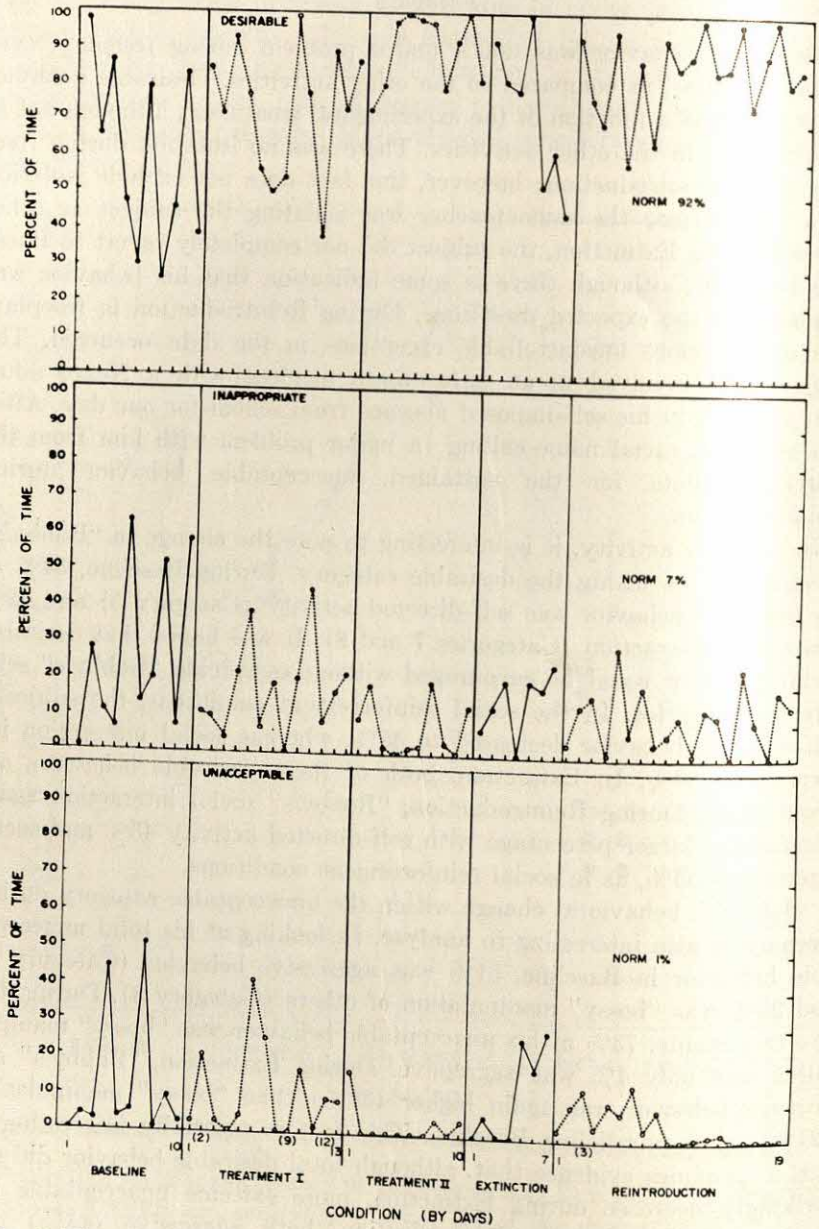


FIG. 2. Discussion: Daily percentage of desirable, inappropriate, and unacceptable behavior by experimental condition; days during which isolation occurred are circled.

maintenance of the treatment situation and "Bobby's" behavior. In summary, the first check revealed that his behavior during freeplay was 99% desirable and 1% unacceptable. At the second check, his behavior was 89% desirable, 1% inappropriate, and 10% unacceptable. Although the latter estimate of unacceptable behavior is higher than expected, considering these two days as a sample of post-experimental behavior, it is evident that no gross or consistent reversion had occurred in freeplay.

Discussion

"Bobby's" behavior in discussion was extremely variable during the Baseline as shown in Fig. 2. In Figs. 2 and 3, for those days in which isolation occurred, the percentages are based on total time excluding isolation time. The extent of class disruption by "Bobby" can be evaluated by comparing his behavior to the norm for boys (1% unacceptable behavior) in the same kindergarten. In Treatment I, and more strikingly in Treatment II, there was improvement in "Bobby's" behavior and an increase in stability.

"Bobby" was isolated for the first time on the second day of Treatment I during discussion. He did not remain in isolation as directed and, therefore, was taken to the principal's office and left for 10 minutes. After this occurrence, "Bobby" responded to the teacher's warning and was not isolated again during discussion in that treatment. As shown in Figs. 2 and 3, his behavior in discussion and rest still needed improvement and, therefore, the stricter second treatment was introduced. On days 1 and 2, "Bobby" was isolated again in line with the more stringent contingencies, but did not require being taken to the principal's office. Thus, in Treatment II, unacceptable behavior decreased strikingly and inappropriate behavior became more consistently lower. In Extinction, one notes a definite deceleration in "Bobby's" desirable behavior. Again, in Reintroduction his behavior improved, although somewhat erratically.

In looking at "Bobby's" unacceptable behavior during discussion, his negative attention-getting behavior (Category 2) was the highest percentage throughout the conditions. Quite often, "Bobby" appeared to be bored by the discussion and would mimic other children, make loud noises, or annoy others.

During the two maintenance checks after Reintroduction, "Bobby's" behavior in discussion was maintained at a high desirable level, 99% after 2 weeks and 95% after 4 weeks.

Rest

"Bobby's" behavior was the most variable during rest (Fig. 3) in the Baseline condition and at times rose greatly above the norm of 5% for

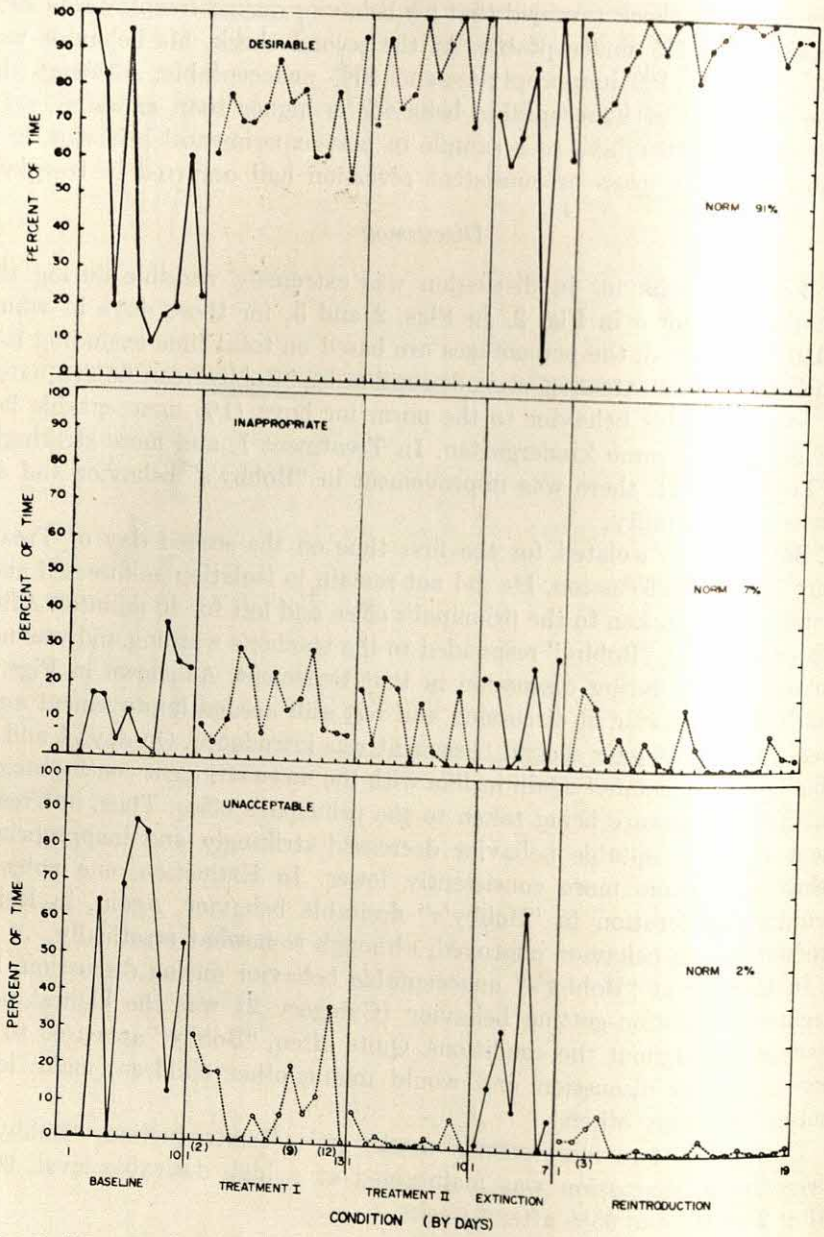


FIG. 3. Rest: Daily percentage of desirable, inappropriate, and unacceptable behavior by experimental condition; days during which isolation occurred are circled.

unacceptable behavior. With the introduction of the first treatment his unacceptable behavior became less extreme, but it was still too high and inappropriate behavior did not appear sufficiently affected. When the stricter second treatment was applied, his inappropriate behavior decreased over-all when punished, while desirable continued to increase and unacceptable to decrease.

During the two treatments of social reinforcement, "Bobby" was isolated five times during rest. On the second day of first treatment, when he had previously been isolated and sent to the principal's office during discussion, he was isolated again in rest, and he stayed there quietly. In the first days after Christmas vacation, "Bobby" was isolated twice. Presumably, he was testing the limits again. On the first day of the second treatment, he was also isolated and at one other time during Treatment II. In Extinction, "Bobby's" behavior became less desirable and the unacceptable increased strikingly. In Reintroduction, his behavior improved rapidly and decisively. It is interesting to note that the experimental treatment appears to have been most effective in the situation in which "Bobby" was most unpredictable and most extremely a problem.

In rest during the two maintenance checks, "Bobby" maintained a high desirable level, 100% after two weeks, 91% after 4 weeks.

DISCUSSION

The baseline data supported the teacher's concern about "Bobby's" disruptive, resistant, and aggressive behaviors. They also indicated that his behavior was extremely variable and unpredictable. With the introduction of the first treatment of positive social attention contingent upon desirable behavior, ignoring all inappropriate behavior, and punishing unacceptable behavior, "Bobby's" behavior showed some improvement. With the second treatment of giving positive social attention contingent upon desirable behavior and punishing inappropriate and unacceptable behavior, his behavior stabilized at a satisfactory level. During Extinction, "Bobby's" behavior regressed sufficiently to indicate that the reinforcing teacher's attention and isolation from attention were significant independent variables affecting his behavior. In the Reintroduction of experimental Treatment II, the subject's behavior returned to a desirable level. "Bobby" was no longer considered a behavior problem by the teachers.

The presence of two other teachers in the kindergarten somewhat complicated the administration of reinforcement by the experimental teacher. The others were at times reticent to interact with "Bobby" or

sometimes they would inadvertently contradict the reinforcing teacher. As a result, subsequent studies have involved collecting data on all teachers present. The method of analyzing teacher interactions requires improvement and consideration of the reliability of frequency of interaction.

After a survey of literature in this field, this investigation appears to be one of the first (Gallagher, 1967) to employ a comprehensive system of categorization to measure modifications of behavior. Previously, operant studies have discussed a single behavior to be analyzed and modified. In the present study, the continuous and all-inclusive teacher interactions give a comprehensive picture of the situation. In addition, few other studies of this nature have attempted to analyze the reinforcing agent's actual behavior after a general description of the proposed treatment. Such careful observations have been useful as a teacher-training device. As a by-product, a systematic technique of developing verbal disapproval as a conditioned punishing stimulus has been employed with other behavior problems.

The results of this study indicate that the systematic use of social reinforcement techniques in the classroom can significantly change a child's behavior, even when the target is more comprehensive than a single operant. The procedures described offer a clear, objective guide for discriminating occasions to present and to withhold positive social reinforcement. It is encouraging to note that behavior modification techniques are applicable to some of the behavioral problems presented by the "disadvantaged" child to increase his learning potential in the classroom.

REFERENCES

ALLEN, K., HART, B., BUELL, J., HARRIS, F., AND WOLF, M. Effects of social reinforcement on isolate behavior of a nursery school child. In Ullman, L. P. and Krasner, L. (Eds.), *Case studies in behavior modification*. New York: Holt, 1965, 307–312.

BRISON, D. Case studies in school psychology: A non-talking child in kindergarten. Urbana, Illinois: University of Illinois, 1966. (mimeographed paper)

GALLAGHER, J. Classroom behavior modification techniques applied to educationally deprived, primary children. Durham, N. C.: Education improvement program, Duke University, 1967. (mimeographed paper)

HARRIS, F., JOHNSTON, M., KELLEY, C., AND WOLF, M. Effects of positive social reinforcement on regressed crawling of a nursery school child. *Journal of Educational Psychology*, 1964, **55**, 35–41.

JOHNSTON, M., KELLEY, C., HARRIS, F., AND WOLF, M. An application of reinforcement principles to development of motor skills of a young child. *Child Development*, 1966, **37**, 379–387.

SPAULDING, R. A coping analysis schedule for educational settings (CASES). Dur-

ham, North Carolina: Education Improvement Program, Duke University, © 1967.

STAATS, A. (Ed.) *Human learning.* New York: Holt, 1964.

ULLMAN, L., AND KRASNER, L. (Eds.) *Case studies in behavior modification.* New York: Holt, 1965.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 220–230 (1969)

Spelling Performance of Left-Handed Schoolchildren as Affected by the Use of a Pencil Modified to Increase Visual Feedback

DOROTHY URE[1]

Department of Psychology, University of Aberdeen, Scotland

Educationists have noted that left handedness in certain cases seems to constitute a handicap in scholastic performance although reasons for this are not altogether clear. Difficulties in spelling, reading, and writing are usually attributed to the inadequate development of cerebral dominance, crossed laterality, and so on. It is suggested here that, without invoking such factors, poor spelling for example, may be due to the fact that the left hander, moving from left to right, tends to obscure the word as he writes it, thus preventing himself from seeing and developing the shape or gestalt of the word—a most important cue in spelling. The experimental pencil was designed to allow this visual feedback, while writing in the normal position, and field trials with 44 left handers in appropriately controlled conditions showed a significant improvement in spelling even in a one-test trial. Improvement was due largely to correction of reversals in the younger groups, and of random errors of omission in the older groups.

Left-handed writers frequently adopt compensatory postures in an attempt to facilitate visual feedback—for example turning the page round through almost 90°, then writing from top to bottom instead of from left to right—or alternatively curling the fist up and over the line so that the word is written from above downwards. Since this reverses or inverts the normal push/pull axis of the hand to the body, it seems likely that it would foster errors of reversal or inversion, particularly in young children who, in the early days of writing, are guided by muscle cues as much as by visual analysis.

In "The Backward Child" Professor Burt says of the left-handed child, "His paradoxical task is to produce with the left hand, a style of writing evolved for the right," and this simple mechanical difficulty is sometimes the source of a variety of unfortunate consequences for the left-handed child.

Margaret Clark has analyzed a number of these in her book "Left Handedness" (1957). For example, the right-handed writer moving from

[1] For their kindness and co-operation during the field trials of this pencil, I am indebted to the late Miss Gladys Bruce, Miss M. Reid, Mr. W. C. Marwick, M.A., and, in particular, to Mr. John M. Wright, M.A.

left to right, pulls his pencil after his hand as he writes. The left-hander, to move in the same direction, must push his, and in so doing, is liable to spike and plough up his page if his pencil is hard and sharp, or, if it is soft, smear what he has just written as his hand passes over it.

It is singularly unfortunate for the left-handed child in those first few months of school life, when attitudes are being formed that may influence the rest of his scholastic life, that so many unhappy repercussions are liable to attend the fact that his writing is likely to be (a) "dirty" and (b) "falling over backwards"—both faults that are frequently denounced by a teacher who would not dream of scolding a child who was unable to read or count.

Another unfortunate effect, however, which does not seem to have received any consideration, is that the left-handed child, unless he is a contortionist (which some are) tends to obscure the series of letters he is forming as his hand moves over them. In other words, he is deprived of the visual cue of seeing the word develop as he writes it.

That this visual feedback is important to him, can be gauged by the expedients the left-hander may resort to in order to facilitate it. (See Fig. 1 for commonest writing postures adopted by left-handed writers).

Moreover, the visual gestalt of the total word is one of the most important cues in spelling (Schonell, 1945). Most adults, even, have had the experience, when doubtful over the spelling of a word, of writing down the two alternative versions and studying them, in order to see which "looks" right, before being able to decide.

With regard to the left-hander's traditional liability to reversals, it is noteworthy that Professor Burt who studied a group of mirror writers, found them to be predominantly left-handed, and attributed their tendency to reversals, to the fact that their writing was uncontrolled by vision but seemed to be guided entirely by kinesthetic or muscle cues. He suggests that young children (and also a number of older children who are poor performers) are "motor minded" rather than "eye minded" to begin with, and that a great step forward is taken when visual control of the writing is fused with, and then gradually takes over from, the primitive muscle cues. When this happens, the reversals so characteristic of the writing of young children tend to die out, and the way is prepared for a great step forward in speed and accuracy.

It seems possible that the simple, mechanical fact of being unable to see around a corner may account for the unexplained shortcomings of many left-handers, by hindering them from making this vital switch to dependence on visual control and analysis.

In extreme cases of reading and spelling disability amounting almost to dyslexia, it has been noted more than once, that the ability to preserve

the order of a sequence is peculiarly weak (Engler, 1917; Laubenthal, 1941; Bender, 1956). A good, inborn ability to maintain a sequence will compensate for, and offset, this lack of the visual cue, leaving both teacher and pupil unaware of any handicap, but if a child with a rather poor short-term memory for letter sequence happens also to be a left-hander who obscures each letter of the word as soon as he has laboriously formed it, one can see how he would be doubly handicapped thereby, not only in the writing of spelling, but also in the learning of it—which is most commonly done by writing. In fact his disability would have cumulative effects of mislearning and failure.

The subject of left-handedness, incomplete or mixed laterality and cerebral dominance, is, of course, a most complex one, and the interacting effects of mechanical, psychological, genetic, and functional factors complicate the picture of the left-hander still further. The present experiment, however, attempts to isolate from this very complex matrix, only the simple factor of impaired visual feedback, and its specific, one-trial effects on spelling performance as measured in school.

The children who performed this test were already categorized as good or bad spellers, with well-established engrams of familiar words already in their minds. Obviously, the increase of visual feedback in one test trial would be unlikely to have very striking results or to alter established misconceptions either. Only over a period of learning could it be hoped that the use of the experimental pencil might improve performance substantially, and then only in appropriate cases. However, by using large numbers, in a closely controlled situation, it was hoped that trends in the appropriate direction might at least point the way for a long-term study.

Figure 1 illustrates the most common postures adopted by left-handed writers. It will be noted that the deviations from the normal (right-handed) position are designed for, or at least result in, an improved view of the word as it is being written. The commonest, Fig. 1a, makes the writing somewhat more visible by hooking the arm and wrist above the line and writing from above downwards. The word being written is revealed to a greater or lesser extent, but it will be noted that the subject is, in fact, writing upside down in the sense that the normal push/pull axis of the hand to the body is reversed.

It is worth noting that the adoption of such a manner of writing would lay a child more than usually open to mistakes in the form of reversals and inversions of letters and word parts.

The second method, Fig. 1b, sometimes referred to by teachers as "Chinese writing," involves turning the paper sideways and writing from above downward towards the body, i.e., from top to bottom, instead of from left to right. The visibility of the word being written is thereby

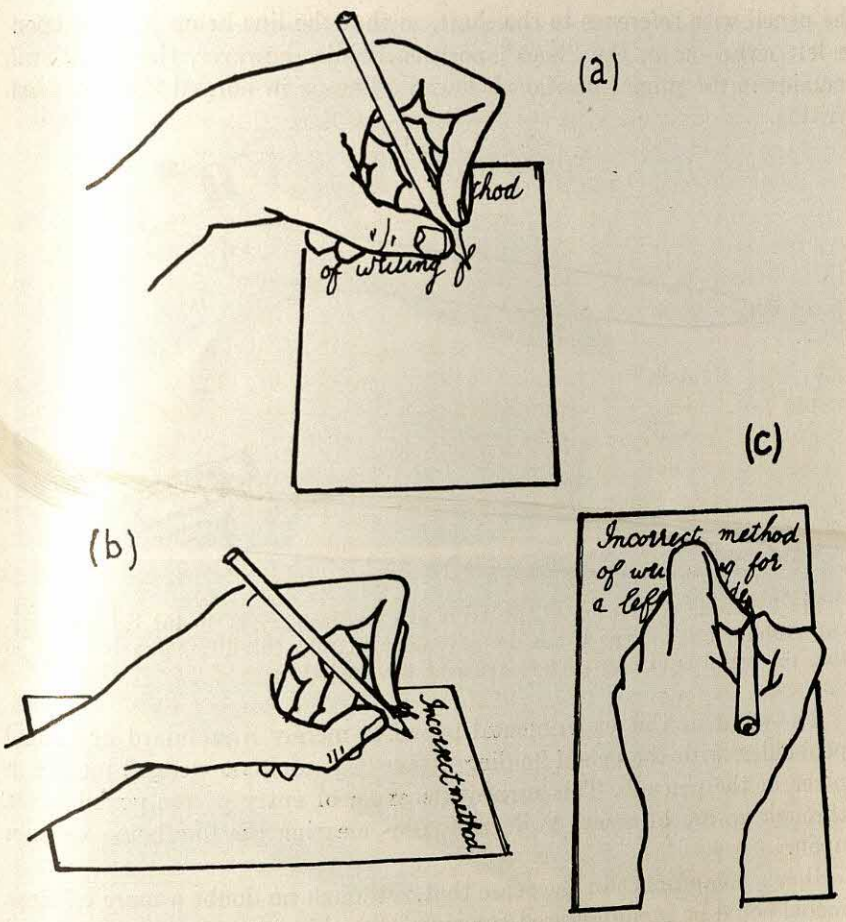


FIG. 1. Illustration of incorrect writing positions adopted by left-handed writers: (a) arm hooked above writing; (b) writing in toward body; (c) with arm cramped into side. Reproduced by permission of the Scottish Council for Research in Education, from "Left-Handedness" by Margaret M. Clark.

increased, but its intelligibility is, of course, reduced, and this sideways rotation of the normal axis of perception would also appear to promote reversals and confusions.

In the normal (i.e., as for right-handers) position, Fig. 1c, the word being written is obscured by the fingers holding the pencil.

THE EXPERIMENTAL PENCIL

The experimental pencil, shown in Figs. 2 and 3, is an attempt to obviate as many of these difficulties as possible, by angling the point of

the pencil with reference to the shaft, so that the line being written upon is left clear—as in the "hook" position, while, however, the hand still remains in the proper position below the line as in normal right-handed writing.

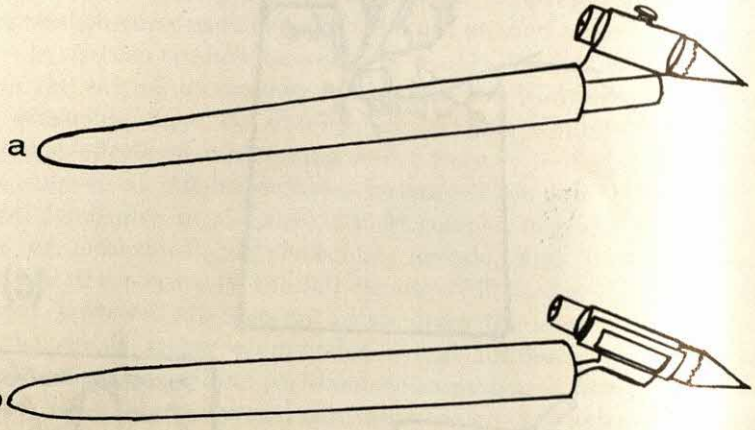


FIG. 2. Schematic diagram of two types of experimental pencil. (a) Screw fitting. (b) Push-in fitting. Type b has the advantage in that the fitting can be bent so that the pencil is at any angle convenient for the writer.

The shaft of this experimental pencil is merely a standard or school penholder, with the pencil-holding fitting of a compass wedged into it in place of the pen nib. This turned the angle of entry of the pencil point through an arc of about 45 degrees, thus clearing the line being written upon.

It is a compromise in the sense that, although no doubt a more efficient model could be manufactured commercially, this prototype was produced easily, quickly, and cheaply, and refilled indefinitely with ordinary pencil stubs. Most of the compasses used in the experiment cost ninepence from a chain store, and had a simple push-in type pencil fitting. Expensive brass compasses with large screw fittings for the pencil tended to obscure the view, and defeat the purpose of the modification. The extra area to the left of the pencil point, revealed by the use of the experimental pencil, is not great—rather less than an inch, but half an inch here is crucial, for it is about the length of an average word, and enables the child to perceive and develop the shape, or gestalt, of the word as he writes it.

Moreover, classrooms are designed and seating arranged with the right-handed writer in mind, i.e., with the light coming in from the left-hand side. The left-handed child, who does not cover the word as he writes

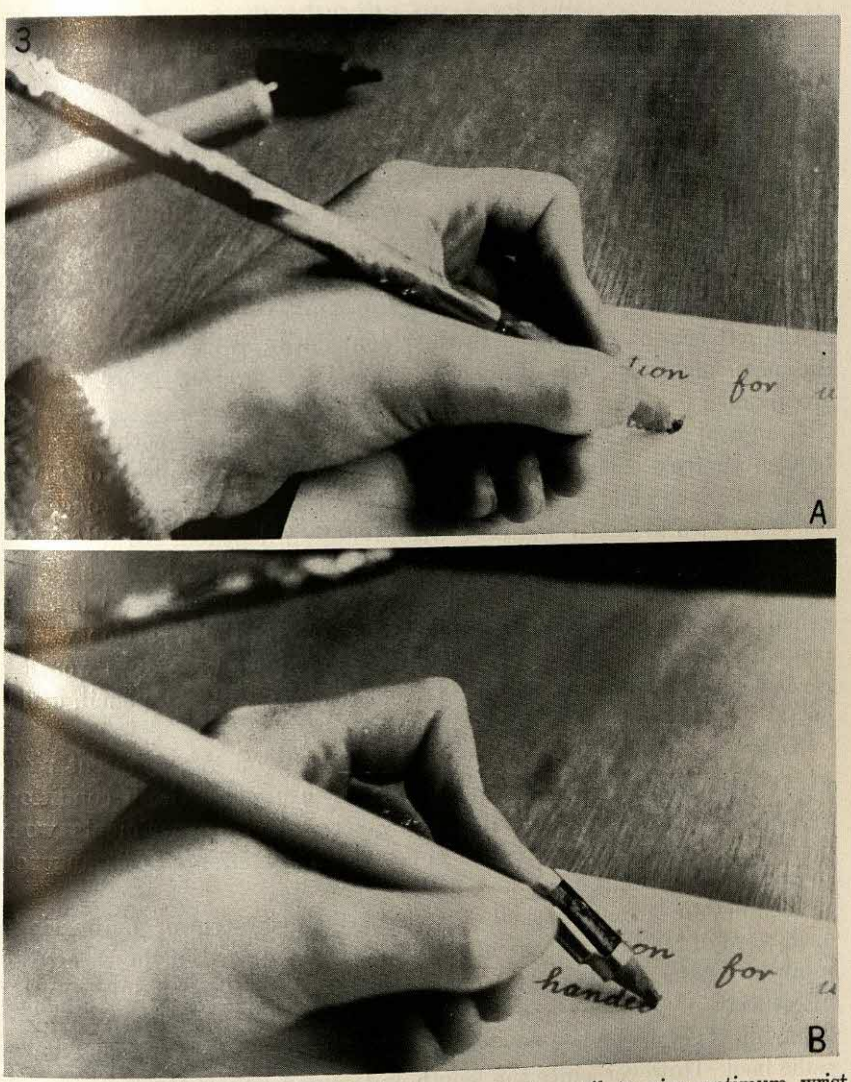


FIG. 3. A. Recommended position with standard pencil to give optimum wrist movement and visibility. [Clark (1957); Cole (1939); Gardner (1945).] B. Wrist freedom and normal position of hand and paper maintained with experimental pencil—and crucial word area revealed. (British Provisional Patent No. 51409/68.)

it, certainly shades it, and, in peering under his hand against the light, is unlikely to see very much. The experimental pencil may be presumed at least to ease matters by lowering the hand to a normal position below the line where it does not shadow the writing.

SUBJECTS AND PROCEDURE

The left-handed subjects of the experiment, 44 in all, came from two infant and two primary corporation schools, and ranged in age from top infants through all grades of primary. They were identified as left-handed writers by their teachers and each left-hander was accompanied to the test classroom by two right-handed counterparts, selected by their teacher as being, as far as could be judged, equal in spelling ability and performance to the left-handed subject. The selection of right-handed controls was made solely on spelling performance levels—other attainment levels being disregarded. Subsequent comparison of the performance of the left-handed subjects with their controls indicated how very good this assessment was, the difference between the average score for each group under the same conditions being insignificant.

Groups of about 24 children, that is, eight left-handers with two right-handed controls each, were tested at one time with a piece of unseen dictation. Each child wrote the same piece of dictation twice—once with the experimental pencil and once with a standard pencil. An alternating schedule was arranged, so that at one session the left-handers started with the experimental pencil, while the right-handers used an ordinary pencil. On completing this they changed over and the dictation was repeated. However, in the next group to be tested the procedure would be reversed, so that the left-handers started with the ordinary pencil, and so on.

No comparison of performance could be made on two different pieces of dictation, so the same test piece had to be used for both conditions, but it was hoped that this alternating schedule of experiments would successfully balance out practice effects and any effects of strangeness, fatigue, or boredom.

The dictations were marked on a simple point system of one mark for each word of text correctly rendered. The material used in the case of the infant departments was Test A of Graded Dictation Tests from Schonell's "Essentials of Teaching and Testing Spelling" while that of the primary departments was Review 40 from the Graded Dictation Reviews in the same book. However, since the length and complexity of the test piece for the primary departments had to be arbitrarily adjusted or curtailed to suit the various attainment levels of the groups as they were encountered, the dictation was not recorded but spoken. Although every care was taken to make the two renderings as nearly identical as possible for pace, distinctness, and/or inflection, it still left the possibility that the experimenter might involuntarily speak more distinctly during those trials when the experimental pens were in use by the left-handers—thus perhaps improving their performance. This would, of course, also mani-

fest itself as a corresponding improvement in the performance of the right-handers who were meantime using ordinary pencils, but since this eventuality might be expected in any case, it left the possibility of experimenter bias still uncertain. To solve this, the second right-handed control in each case was used to form a simple replication group. They used ordinary pencils for both dictations, and their improvement or deterioration could be expected to reflect any experimenter bias that might be operating.

RESULTS

In fact, the evidence from this "no treatment" group showed either no difference at all from one dictation to the next, or a marginal improvement during that dictation when the left-handers were *not* using the experimental pencils—indicating, if anything at all—a tendency to overcompensate for experimenter bias.

TABLE 1

SPELLING SCORES ACHIEVED BY GROUPS OF LEFT-HANDERS USING STANDARD AND EXPERIMENTAL PENCILS ALTERNATELY[a]

School group	Subject	Scores using standard pencil	Scores using experimental pencil
Infant	1	7	7
	2	10	11
	3	11	12
	4	8	8
	5	7	7
	6	6	8
	7	4	4
	8	9	10
Total	8	62	67
Infant	25	5	9
	26	10	10
	27	12	12
	28	12	12
	29	12	13
	30	11	12
Total	6	62	68
Primary	43	17	15
	44	29	29
	45	22	23
	46	12	17
	47	29	29
	48	28	28
	49	29	29
	50	29	29
Total	8	195	199

TABLE 1 (*Continued*)

School group	Subject	Scores using standard pencil	Scores using experimental pencil
Primary I and II	67	15	14
	68	15	17
	69	13	13
	70	8	10
	71	21	22
	72	20	20
	73	24	24
Total	7	116	120
Primary II and IV	88	26	26
	89	28	28
	90	27	27
	91	26	26
	92	28	28
	93	23	23
	94	25	25
Total	7	183	183
Primary V and VI	109	21	22
	110	28	28
	111	28	28
	112	24	25
	113	28	28
	114	27	27
	115	25	26
	116	27	27
Total	8	208	211
Over-all total	44	826	848

[a] MD = .5; df 43; significant at .01 level.

The right-handed control group who were tested on both pencils also showed no significant difference in performance or, if anything, a slight deterioration in some cases with the experimental pencil—the result no doubt of strangeness or awkwardness effects.

The left-handers, however, in spite of having to adopt not only a new kind of pencil, but also in many cases a new kind of grip, appeared to benefit significantly over-all from the use of the experimental pencil (Table 1). Some subjects, notably the high achievers, seemed relatively unaffected, while others benefited quite substantially. The improvement (significant at the .01 level using the t test, and at the .001 level using the Sign Test and the Null Hypothesis for the binomal test $p = q = \frac{1}{2}$) seemed to come mainly from a reduction of reversal-type errors among the infant pupils, and of random mistakes in older pupils, particularly a cutting down of errors of omission.

DISCUSSION

As can be seen from Fig. 4, established word concepts did persist from one trial to the next, particularly those formed on an auditory basis, e.g., "cins" (kinds), "prity" (pretty), "erly" (early), "wota" (water).

Improvement in the infant classes came mostly from the correct rendering with the experimental pencil of common reversals of letters, such as "b" for "d" in "deep," "d" for "b" in "blue," and word parts, such as "sae" to "sea."

With the primary school children the number of manifest reversals was much smaller, except in the case of one or two notably bad spellers, and improvement effects came mostly from the prevention of errors of omis-

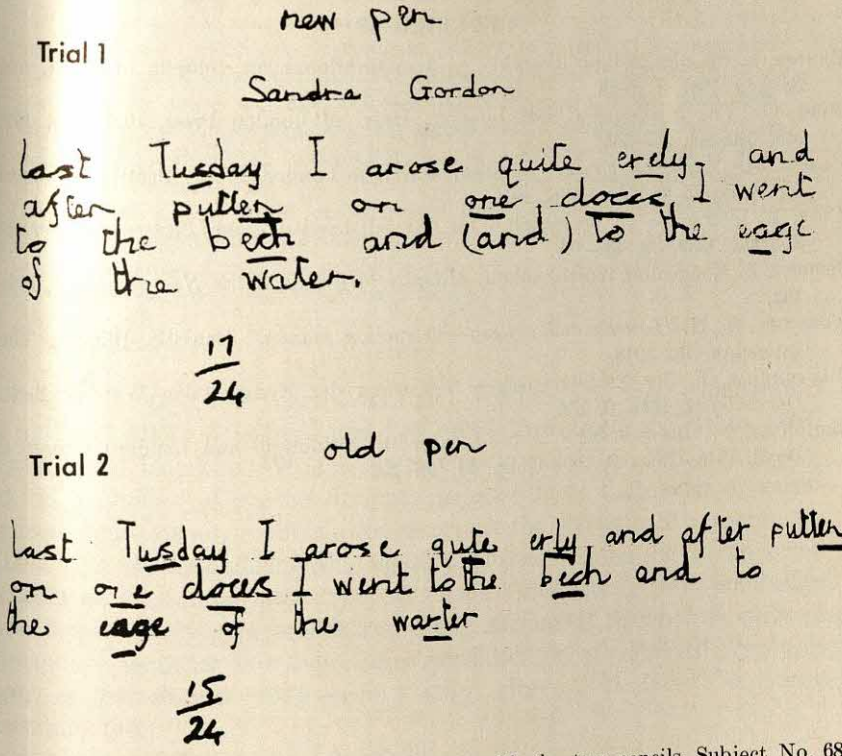


FIG. 4. Specimen of compared performance with the two pencils. Subject No. 68, Trial 1 (experimental pencil); Trial 2 (standard pencil). Note repetition of established engrams, especially those formed on an auditory basis, e.g., "Tusday," "putten," "eage," "one closes" (warm clothes). Improvement effect of the new pen due to correction of omission in "quite" and a correct visual analysis of "water." These "improvement" effects were of course, in the first trial, and reversion to the normal pencil in Trial 2 showed deterioration instead of the practice effects one would have expected from the corresponding performance of the "no treatment" control group, whose score improved from 127 to 132 in Trial 2.

sion such as "menbers" (members), "qute" (quite), and odd mistakes such as "clother" (clothes).

However, in a number of cases of peculiarly bad spelling, quite a few of the words had changed with the experimental pencil to a more logical approximation of the word, and, although still wrong, had lost the peculiar cross-grained quality that characterized them, e.g., "closth" to "closes."

Improvement was for the most part marginal in individual cases, but the fact that it established itself stably in a one-test trial in spite of the "flattening" effect of previous learning, suggests that cumulatively beneficial effects might be expected from the early use of the pencil in appropriate cases.

REFERENCES

BENDER, L. Specific reading disability as a maturational lag. *Bulletin of the Orton Society*, 1957, **7**, 9–18.

BURT, C. *"The Backward Child"* London: Univ. of London Press, 1937. Pp. 281, 318, 340–344, 353–355.

CLARK, M. *"Left-Handedness"* London: Univ. of London Press; Scottish Council for Research in Education, 1957.

COLE, L. Instruction in penmanship for the left-handed child. *Elementary School Journal*, 1939, **39**, 436–448.

ENGLER, B. Kongenital Wortblindheit. *Monatsschrift Psychiatrie Neurologie* **42**, 119–132.

GARDNER, W. H. *"Left-handed writing—instruction manual"* Danville, Illinois: The Interstate Co., 1945.

LAUBENTHAL, F. Zur Erbhygienischen Bewartung der Kongenitalen Wortblindheit. *Der Erbartz*, 1941, **9**, 156.

SCHONELL, F. *"Backwardness in basic subjects"* Edinburgh and London: Oliver & Boyd, 1945. Chap. 9, 109; Chap. 14, 297–298.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 231–244 (1969)

Intensity of Punishment, Timing of Punishment, and Cognitive Structure as Determinants of Response Inhibition[1]

J. ALLAN CHEYNE AND RICHARD H. WALTERS
University of Waterloo

Eighty-four Grade 1 boys were randomly assigned to 1 of 12 experimental or two control conditions. Experimental subjects were trained under one of six punishment conditions, varying in respect to intensity of punishment, timing of punishment, and cognitive structure. Results indicated that when cognitive structure was low, response inhibition was stronger after early, than after late punishment and that high-intensity punishment was a more effective inhibitor than low-intensity punishment. The provision of high-cognitive structure was very effective for inducing response inhibition, especially when punishment was delivered late. Telemetered heart-rate data secured throughout the experiment further support the suggestion that, under certain conditions, response inhibition may be produced primarily by emotional arousal, and, under other conditions, by the availability of prohibition rules.

During the past few years, Aronfreed, Walters, and their collaborators (Aronfreed, 1965; Aronfreed and Leff, 1963; Aronfreed and Reber, 1965; Walters and Demkow, 1963; Walters, Parke, and Cane, 1965) have conducted a number of studies designed to investigate the effects of variations in punishment training procedures on the strength of response inhibition in children. Generally speaking, these studies have supported hypotheses that high-intensity punishment produces greater inhibition than low-intensity punishment and that punishment that occurs early in a response sequence is a more effective inhibitor of punished responses than is punishment that occurs only after a response has been consummated.

[1] The study reported in this paper was supported by grants to R. H. Walters from the Ontario Mental Health Foundation (No. 42), the National Institute of Health Service (HD 01456-02), the Defence Research Board of Canada (No. 9401-24), and the National Research Council of Canada (No. APA-94). The authors acknowledge the advice of Dr. P. C. Jennings on statistical analysis and the assistance of Peter Birrel and Philip Zelazo in the collection of the data. The co-operation of the Superintendent of the Galt School Board and of the Principals and Staff of Central, Highland, and Stewart Avenue Public Schools is gratefully acknowledged.

In contrast, Parke and Walters (1967), in a more recent series of studies, found that low-intensity punishment paired with a consummatory response was as effective for inducing response inhibition as high-intensity punishment delivered at the commencement of a deviant act. This finding is difficult to reconcile with two-factor learning model (Mowrer, 1960a, 1960b) on which predictions were largely based. Low-intensity punishment delivered late in a response sequence would result in the association of little, if any, anxiety with the preparatory and initiating responses; consequently, one would expect the effects of anxiety to be counteracted, at least to some extent, by secondary reinforcers associated with the consummatory response.

In seeking an explanation of this somewhat paradoxical finding it may be suggested that the low-intensity punisher, a 50 db. auditory stimulus, may have had primarily the function of eliciting an orienting response (e.g., Sokolov, 1963), when delivered late in a response sequence and may thus have served to facilitate the reception of the verbal information conveyed by the comment of the experimenter, "No, that's for another boy," that immediately followed the presentation of the physical stimulus. Thus, a condition was perhaps created under which a verbal instruction, making reference to a *rule* of conduct, exerted considerable control over the behavior of the subjects under the low-intensity—late-punishment condition.

An additional *post-hoc* analysis of data from the Parke and Walters study (Cheyne, 1966) lent support to a hypothesis, that over a short testing period there would be differences between subjects whose inhibition was mediated primarily by anxiety and those whose inhibition was mediated primarily by the rule. In the case of anxiety-mediated inhibition, deviation in the absence of punishment should lead to a reduction in the anxiety associated with the deviant response, and hence there should be an increase in deviation over time during a continuous testing period. On the other hand, a lack of subsequent surveillance and hence punishment should not change a child's knowledge that the response is, nevertheless, prohibited by some rule. Therefore, to the extent that the child's behavior is controlled by the rule, the amount of deviant responding should not change over time.

The present experiment was designed to determine the effects on response inhibition produced in children, by various punishment techniques varying in the intensity and temporal relation of punishment to a response to a particular object and particularly the way in which the intensity and timing of punishment may be modified by the provision of cognitive structuring techniques which provide the subject with verbal information which explains or justifies to the subject the punishment operations.

In order to further test by another measure the hypotheses that there may be two partially independent mediational processes involved in the learning of self-control—one emotional and one cognitive-telemetered heart-rate data were secured from the subjects throughout the experimental proceedings.

METHOD

Subjects

Eighty-four Grade 1 boys were randomly assigned in equal numbers to twelve experimental and two control groups. The experimental subjects were trained under one of six punishment conditions, varying in respect to timing of punishment (early vs. late) and to the nature of the punishment procedure (presentation of a high-intensity noise alone, of a low-intensity noise alone, or of a low-intensity noise immediately followed by a verbal signal). Half the subjects under all punishment conditions were exposed to a filmed sequence designed to elicit deviant behavior, while the remaining subjects were exposed to a control sequence.

Subjects who received punishment in the form of a low-intensity noise accompanied by a verbal signal were also, before punishment training, provided with reasons why they should not deviate. This procedure may therefore be regarded as providing high-cognitive structure in comparison to the other punishment procedures under which cognitive structure was relatively low.

Experimental Arrangements

The subjects were tested in a mobile laboratory, divided into an experimental and an observation room by a partition containing a one-way vision screen. Doors connected the two rooms and permitted access to either room from outside the laboratory. Punishment training was carried out on a table to which two buzzers, one of 96 db and one of 54 db were attached below the top surface. For the resistance-to-deviation test, subjects were seated at a second table. The films were presented on an Admiral Classroom Television monitor in closed circuit with an Ampex VR-700 videorecorder located in the observation room.

During the entire experiment, the subjects' heart-rates were recorded on a Beckman-Offner dynograph by means of a telemetry system. Electrodes were connected to a miniature (1½ inches × 1½ inches × ½ inches) FM receiver to a Beckman-Offner dynograph. Both the receiver and the dynograph were located in the observation room.

Procedure

The children were brought individually by the experimenter from the classroom to the laboratory. Immediately upon a child's entering the ex-

perimental room, the transmitter electrodes were applied about two inches apart, to his sternum.

Punishment training. The subject was first seated at a table and instructed that he would be presented with toys in pairs and that he was to select one of each pair.

Variations in cognitive structure. For subjects assigned to the low-cognitive-structure conditions, the following instructions were given: "Now, some of these toys you should not touch or play with, and if you choose one of the toys you are not supposed to touch or play with, you will know because you will hear this buzzer (demonstrating either the high-intensity or low-intensity buzzer, depending upon the condition to which the subject had been assigned). Do you understand that? O.K."

Subjects assigned to the high-cognitive-structure condition were instructed as follows: "Now, some of these toys you should not touch or play with because I don't have any others like them and if they were to get broken or worn out from boys playing with them, I wouldn't be able to use them any more, so for that reason I don't want you to touch or play with some of these toys. And if you choose one of the toys you're not supposed to touch or play with, I'll tell you and you will hear a sound like this (demonstrating the low-intensity buzzer)."

After these instructions had been given, punishment training began. On five of the nine trials the subjects were punished regardless of which toy they chose. The punished trials were 1, 3, 4, 7, and 9.

Variations in the timing and nature of the punishment procedures. For subjects under the early-punishment condition, the experimenter administered the appropriate punishment as soon as the subject's hand began to move toward one of the toys. If the subject was in a high-cognitive-structure group, the buzzer was followed by the remark, "No, not that one." For subjects in the low-cognitive-structure groups, the sound of the buzzer was not followed by a verbal remark. Subjects under the high-intensity conditions were punished by a 96 db noise, those under the low-intensity conditions by a 54 db noise.

Under the late-punishment condition, the subject was permitted to pick up the toy and hold it for approximately three seconds before punishment was delivered. Again, the sound of the buzzer was followed by a remark only for subjects under the high-cognitive-structure condition.

Response-inhibition test. On completion of the punishment trials, the experimenter placed the punished toys on a second table in two rows (three in the first row and two in the second) behind a screen blocking the subject's view. The experimenter then said to the subject: "Oh, now I have to leave for a while. I have to take something from the other room into the school. Will you wait for me here? Fine. Maybe you would like

to watch television while I'm gone. I'll turn it on, and we'll see if something is on now." The experimenter then removed the screen, revealing the table with the toys and television set. The experimenter turned on the television set and continued: "Well, it will take a few minutes to warm up. Maybe you will be able to see the television better from over here." The experimenter seated the subject on a chair behind the table containing the prohibited toys. He then put on a coat as if in preparation to leave the laboratory, saying: "This door (indicating a door leading outside) is locked so that nobody can bother you while I'm gone. I have to get something from the other room so I'll go out by the other door. Remember about these toys. Goodbye." The experimenter left the room and immediately turned on the videorecorder, thus projecting on the monitor either a sequence showing a Grade 1 boy, who was unknown to the subjects, playing with all possible prohibited toys, each for approximately the same length of time, or a control sequence depicting inanimate objects unrelated to the experiment. He loudly slammed the door leading from the observation room to the outside of the trailer, as if he were leaving; in fact, he remained in the observation room with his assistant.

Response-to-deviation test. After fifteen minutes during which the subject was alone with the toys, the experimenter again slammed the door leading from the observation room, then re-entered the experimental room through the inside door. He offered the subject several candies and asked: "Did you play with any of the toys while I was gone?" The subject was then returned to his classroom.

Behavioral measures. An observer behind the one-way vision screen in the darkened observation room recorded on a record sheet, similar to that used by Parke and Walters (1967), the times at which the subject touched and ceased to touch individual toys. The sheet was arranged in such a way that the observer had only to record the times, read from a Heuer-Century stopwatch, in appropriate squares.

From these records the following scores were calculated: the latency of the subject's first deviant response, the number of times he deviated, and the total time for which he deviated. These scores represent only slight modifications of those used in earlier studies (Walters, Leat, and Mezei, 1963; Walters and Parke, 1964; Walters, Parke, and Cane, 1965; Parke and Walters, 1967). In addition to the above scores, a fourth score was obtained by dividing a subject's duration-of-deviation score by the number of times he deviated. It was also possible to obtain from the observer's records changes in number, duration and average duration per deviation over time. These scores were obtained separately for the first five minutes, the second five minutes, and the last five minutes of the testing period.

In addition to the above records, some supplementary information was secured by the observer. After the experiment had begun and two subjects had been run in each cell, the experimenter suggested that the observer check any deviation during which the subjects lifted the toy from the table, as opposed to merely touching or handling it on the surface of the table, in order to obtain an index of the degree of caution with which the subject handled the toy. Also, the number of candies taken by the subjects during the response-to-deviation test and their answers to the experimenter's questions concerning deviation were recorded.

Physiological Measures

Inspection of the heart-rate data during the punishment-training phase of the experiment indicated that a punishing event characteristically produced a startle pattern that reflected muscle-potential activity as well as heart-rate changes. The number of these responses was counted for each subject for the five trials on which punishment was given. In view of the occurrence of complex startle patterns, change in heart rate could not be utilized as a measure of emotional response during punishment training.

Change in heart rate was, however, employed as an index of the subjects' emotional reactions to the response-to-deviation test. The average heart rates of individual subjects during the last 20-second nondeviation period preceding the termination of the resistance-to-deviation test was used as a baseline index; this index was subtracted from their average heart rates during the 6 seconds following the placing of the candies on the table.

The remaining tests employing physiological data were *post-hoc* designed to clarify findings concerning specific group differences in behavior; the measures employed and the outcome of comparisons are therefore reported in the discussion of results.

RESULTS

Reliability of Measures

A second observer was present while 15 *S*'s were tested. Pearson product-moment correlation coefficients were used as indices of interrater reliability for the dependent measures of response inhibition. Pearson *r*'s for latency, number-of-deviations, and duration-of-deviations measures were all above .99.

Response-Inhibition Data

The means and *SD*'s for the latency, number-of-deviations, and duration-of-deviations measures are presented in Table 1. In spite of hetero-

TABLE 1

GROUP MEANS AND *SD*s FOR THREE INDICES OF RESPONSE INHIBITION

	DEVIANT MODEL											
	Low cognitive structure								High cognitive structure			
	High intensity				Low intensity				Low intensity			
	Early punishment		Late punishment		Early punishment		Late punishment		Early punishment		Late punishment	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Latency (sec)	573.5	329.2	363.0	381.6	438.3	382.2	97.8	118.3	471.	429.5	561.5	336.0
Number of deviations	5.6	7.0	7.7	8.3	7.2	8.1	10.2	3.3	4.3	5.5	1.8	2.5
Duration of deviations (sec)	79.0	111.9	127.8	147.6	100.8	120.2	204.7	151.9	52.0	77.1	6.2	9.5
	NO MODEL											
Latency (sec)	438.3	326.2	209.7	323.0	89.0	91.0	178.8	323.3	230.0	312.1	394.2	291.8
Number of deviations	3.0	3.1	9.8	7.2	12.8	4.9	12.0	5.5	7.5	6.0	2.3	2.3
Duration of deviations (sec)	250.2	34.0	233.5	232.6	146.8	109.9	269.5	291.5	84.0	68.2	7.8	10.3

	Deviant model control		No model control	
	Mean	*SD*	Mean	*SD*
Latency (sec)	55.5	49.9	49.0	20.8
Number of deviations	23.3	6.8	17.2	5.2
Duration of deviations (sec)	368.3	128.5	346.3	161.5

geneity of variance of some of the measures, analysis-of-variance methods were used since the F test has been shown to be relatively insensitive to even large departures from homogeneity (Young and Veldman, 1963). The analyses were carried out by means of orthogonal weighting coefficients (Edwards, 1960).

Subjects who had observed the deviant model initated their first deviation significantly *later* than subjects who had observed the control film ($F = 4.42$; $p < .05$; $df = 1, 70$). None of the other measures revealed a film-manipulation effect. Subjects receiving high-cognitive-structure instructions deviated less often ($F = 8.31$; $p < .01$; $df = 1, 70$) and for a shorter period of time ($F = 8.38$; $p < .01$; $df = 1, 70$) than subjects receiving low-cognitive-structure instructions, but the latencies of the two groups were not significantly different. Latencies were significantly longer ($F = 4.32$; $p < .05$; $df = 1, 70$), and numbers of deviations fewer ($F = 4.84$; $p < .05$; $df = 1, 70$), for subjects experiencing the more intense, as opposed to those experiencing the less intense, noise under the low-cognitive-structure condition. Moreover, under this condition subjects punished early deviated somewhat less quickly ($F = 3.34$; $p < .10$; $df = 1, 70$) and for shorter durations ($F = 7.52$; $p < .01$; $df = 1, 70$) than subjects punished late. There was no significant timing effect under the high-cognitive-structure condition.

The data for subjects in the two control groups were combined and compared with the combined data for the most deviant group under film-model condition and the most deviant group under the neutral-film condition. The control subjects deviated more often ($F = 13.40$; $p < .005$; $df = 1, 70$) and for a longer period of time ($F = 4.04$; $p < .05$; $df = 1, 70$), though not more quickly, than subjects in comparison groups. Hence, there is evidence that even the relatively ineffective punishment conditions produced some inhibition.

In order to test further the hypothesis that cognitive structuring is more effective for producing response inhibition when punishment is delivered late than when punishment is delivered early, two additional nonorthogonal comparisons were made. These comparisons were guided by theoretical considerations and hence were deemed justifiable (Winer, 1962).

Subjects who received punishment early tended, though nonsignificantly, to deviate less quickly, less often, and for a shorter duration of time than subjects who received punishment late; in contrast, under the high-cognitive-structure condition, subjects who received punishment early demonstrated a marked tendency to deviate more quickly ($F = 6.65$; $p < .01$; $df = 1, 70$), more often ($F = 4.86$; $p < .05$; $df = 1, 70$), and for a longer duration ($F = 13.65$; $p < .001$; $df = 1, 70$) than sub-

jects who received punishment late; thus, the usual effect of timing was essentially reversed when high cognitive structure was provided.

Since the hypothesis concerning changes over time involved only the cognitive-structure variable, data for all subjects in the low-cognitive-structure groups were combined and compared with the combined data for subjects in the high-cognitive-structure groups. A trend analysis using the number measure failed to yield significant results. However, the trend analysis based on the duration measure revealed a significant interaction between cognitive structure and time periods ($F = 5.28$; $df = 2, 140$; $p < .01$). The curve for low-cognitive subjects showed a positive slope whereas the curve for high-cognitive subjects showed a slightly negative slope (Fig. 1).

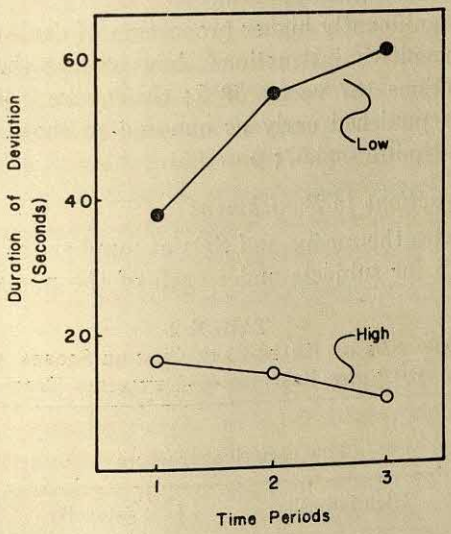


FIG. 1. Change in average duration of deviation over three 5-min. periods for subjects under high-cognitive and low-cognitive conditions.

Additional Response-Inhibition Data

Analyses were carried out for the average-duration-of-deviation measure in the same manner as for the number and duration measures. A trend analysis demonstrated a significant interaction effect involving the cognitive-structure variable and time periods parallel to that obtained with the duration measure ($F = 7.22$; $df = 2, 140$; $p < .001$). Moreover, orthogonal comparisons conducted in order to test between-variable effects showed that subjects who received punishment early in the response sequence deviated for significantly shorter durations per deviation than subjects punished after the response was completed ($F = 6.54$; $p < .02$).

The influence of cognitive structuring on the cautiousness with which toys were handled was treated by comparing deviators who had experienced high cognitive structure and those who had experienced low cognitive structure in respect to the number of times they lifted a toy from the table. Deviators under the high-cognitive-structure condition lifted toys from the table 0.70 times, on the average, compared to an average of 4.15 times for those under the low-cognitive-structure condition ($t = 3.14$; $df = 35$; $p < .01$).

Response-to-Deviation Data

Almost all subjects took only one candy during the response-to-deviation test. This uniformity of behavior probably reflects the influence of a very well established cultural norm.

There was a significantly higher proportion of denials among deviators receiving high-cognitive instructions than among those receiving low-cognitive instructions (87 versus 59%; chi-square = 4.37; $p < .05$) and among deviators punished early as opposed to those punished late (80 versus 41%; chi-square = 4.67; $p = .05$).

Physiological Reactions to Punishment

Table 2 indicates the means and *SD*s of numbers of startle patterns to punishing stimuli for subjects under each of the punishment conditions.

TABLE 2
MEANS AND *SD*s OF HEART-RATE CHANGE SCORES ASSOCIATED WITH THE RESPONSE-TO-DEVIATION TEST[a]

	Low cognitive structure				High cognitive structure	
	High intensity		Low intensity		Low intensity	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Early punishment	13.2	17.2	6.6	4.5	1.5	8.6
	(7.1)[b]	(1.9)	(6.0)	(2.1)	(5.3)	(2.2)
Late punishment	5.3	10.9	4.5	7.9	2.1	6.9
	(4.4)	(1.4)	(4.3)	(2.8)	(4.9)	(1.0)

[a] $N = 8$ in each cell.

[b] Means and *SD*s of Startle Reactions to Punishment.

Since the films were presented only after punishment training was completed, data for subjects under the two film conditions were combined. Heart-rate data were available for 10 subjects with each punishment condition.

Startle patterns were significantly less numerous for subjects under

the high-cognitive-structure condition than for subjects under the low-cognitive-structure conditions ($F = 4.72$; $p < .05$; $df = 1, 53$); however, as Table 2 indicates, this difference was entirely attributable to differences between subjects under the early-punishment condition. Subjects who were punished early under low-cognitive-structure conditions gave significantly more startle responses than subjects punished late ($F = 6.68$; $p < .03$; $df = 1, 53$) the effects of the intensity variable, although in the anticipated direction, failed to reach significance ($F = 1.47$; $p < .10$; $df = 1, 53$).

Table 2 also presents the heart-rate change scores that were used as an index of subjects' emotional responses to deviation. Imperfect recordings reduced the number of subjects with usable records to 57; the records of a further nine subjects were discarded, using randomization principles, to leave data for a total of 48 subjects, eight in each cell. Again, the data for subjects under the two film conditions were combined. Orthogonal comparisons revealed that subjects under the high-cognitive-structure condition displayed less reactivity than those under the low-cognitive structure condition ($F = 5.16$; $p < .05$; $df = 1, 53$); differences attributable to variations in intensity and timing for subjects under the low-cognitive-structure condition approached significance ($F = 3.03$; $p < .10$; and $F = 3.14$; $p < .10$, respectively $df = 1, 53$).

DISCUSSION AND SUPPLEMENTARY DATA

Physiological Data

Since heart-rate data were primarily obtained in order to provide an index of the emotional impact of experimental manipulations on the subjects, it is advantageous to discuss their implications before assessing those of other findings. As Table 2 indicates, timing of punishment was the variable that contributed most to differences among groups in respect to startle patterns changes during punishment. The startle responses that were recorded seem largely to reflect a reaction to a sudden, unexpected intervention by the experimenter as a child was about to make a choice of one of a pair of toys.

The heart-rate change scores associated with the response-to-deviation test (Table 2) provided a slightly different pattern of results. The influence of the timing variable was apparent among subjects under the low-cognitive condition but not among those for whom high-cognitive-structure was provided; in fact, neither early-punished nor late-punished subjects under the high-cognitive-structure condition displayed much emotional reactivity during the response-to-deviation test. At this point, it is worth noting that these indices suggest that the two high-cognitive-structure groups may have differed from the high-intensity—early-

punished—low-cognitive-structure group in respect to emotional responsiveness.

Behavioral Data

Resistance-to-deviation measures. Resistance to deviation was increased by the provision of high-cognitive-structure and, when cognitive structure was low, by the early administration of physical punishment of relatively high intensity.

The hypothesis that response inhibition mediated by emotional arousal is likely to be more weakened over time than inhibition based on the provision of a rule is supported by the data presented in Fig. 1. Especially interesting in this respect is the difference in the degree of "cautiousness" in respect to picking up toys that differentiated high-cognitive-structure children from low-cognitive-structure children. The behavior of high-cognitive-structure subjects appears to have been controlled, at least to some extent, by the specific content of the instructions that they received from the experimenter.

It should be noted that the two least deviant groups of subjects were early-punished subjects who received high-intensity punishment and late-punished subjects for whom cognitive structure was high. Deviant subjects in these two groups were also more likely than any other subjects to deny their deviations. In addition, the physiological data already reported lent support to the suggestion that response inhibition was induced primarily through emotional arousal for early-punishment—high-intensity subjects and primarily through cognitive structuring, accompanied by relatively little emotional arousal, for late-punishment subjects for whom high cognitive structure was provided.

It was hoped that additional relevant physiological data would have been secured during the response-inhibition test. Most of the available data were, however, too contaminated by artifacts resulting from varying activities of the children and, in addition, obviously influenced by a number of inextricable factors, such as the incidence of deviations and variations in length of intervals between successive deviations. Nevertheless, it was decided to make a selected *post-hoc* comparison between subjects under the high-intensity—early-punishment condition and those under the high-cognitive—late-punishment—low-intensity condition in respect to heart-rate changes immediately preceding their first deviation. The average heart rates of each of the six deviators under each of these conditions were calculated for the 24 seconds immediately preceding their first deviation and also for the 5 seconds preceding this deviation. Changes in average heart rates from the earlier to the later of these inter-

vals were used as an index of emotional response to deviation. The median heart-rate change for subjects under the low-cognitive—high-intensity—early-punishment condition was +5.4, while that for subjects under the high-cognitive—late-punishment condition was −0.03. A Mann-Whitney *U*-test for small samples indicated that this difference was significant at the .002 level ($U = 1$; $n_1 = n_2 = 6$). Thus, one might infer that "emotionally-controlled" subjects displayed considerable emotion when deviation occurred, whereas the "cognitively-controlled" subjects showed little emotionality when they broke the prohibition.

This finding would seem to further support the view that the two groups of subjects, although behaviorally very similar, both in respect to their high resistance to deviation and high incidence of denial, were quite dissimilar in their emotional reactions to the extent that heart rate change may be inferred to indicate emotional responses. Their dissimilarity in cardiac activity during the punishment training and during the response-to-deviation test has already been noted. The findings thus suggest the possibility that there are two sources of response inhibition, one generated primarily by emotional responses to anticipated deviation and the other based primarily on knowledge that the certain potential acts are instances of socially disapproved behavior. In the latter case, reference is explicitly or implicitly made to a standard or rule that may be expressed in a proposition; the behavior involved may thus be regarded as an example of rule-regulated or proposition-governed behavior.

Distinctions have frequently been made between fear-controlled and guilt-controlled behavior, on the assumption that there are two discriminable kinds of emotional responses that maintain conformity to standards. Traditionally, fear has been interpreted as an *emotional* response to the actual or anticipated disapproval of others and guilt as an *emotional* response to the actual or anticipated violation of "internalized" standards or rules. We are advocating instead a distinction between emotional and cognitive control. Sometimes, we believe, conforming behavior may result primarily from the activation of conditioned emotional responses by cues that are present when an agent commences to execute, or anticipates executing a socially disapproved act. At other times, we believe, conformity occurs because a potential agent recognizes that a potential act is contrary to a particular rule which, on the basis of past learning has become an effective cue governing behavior.

The proposed distinction does not require the assumption that one source of control is not a developmental antecedent of the other; perhaps "cognitive control," as we have described the phenomenon, could not develop in the absence of some fear-inducing avoidance training. Our find-

ings suggest, however, that the development of self-control is not a matter of learning a "new" *emotional* response that is a substitute for fear but of learning how effectively to utilize socially significant cues.

REFERENCES

ARONFREED, J. Punishment learning and internalization: Some parameters of reinforcement and cognition. Paper read at Biennial Meeting of the Society for Research in Child Development, Minneapolis, March, 1965.

ARONFREED, J., AND LEEF, R. The effects of intensity of punishment and complexity of discrimination upon the generalization of an internalized inhibition. Unpublished manuscript, University of Pennsylvania, 1963.

ARONFREED, J., AND REBER, A. Internalized behavioral suppression and the timing of social punishment. *Journal of Personality and Social Psychology*, 1965 **1**, 3–16.

BANDURA, A., AND WALTERS, R. H. *Social Learning and Personality Development.* Holt, New York, 1963.

CHEYNE, J. A. The influence of intensity and timing of punishment and cognitive structure on the effectiveness of punishment training for inducing response inhibition. Unpublished M.A. thesis, Univ. of Waterloo, 1966.

EDWARDS, A. L. *Experimental Design in Psychological Research.* Holt, New York, 1960.

MOWRER, O. H. *Learning Theory and Behaviour.* New York: Wiley, 1960. (a)

MOWRER, O. H. *Learning Theory and the Symbolic Processes.* New York, Wiley, 1960. (b)

PARKE, R. D., AND WALTERS, R. H. Some factors influencing the efficacy of punishment training for inducing response inhibition. *Monographs of the Society for Research in Child Development*, 1967, **32** (1), (Serial No. 109).

SIEGEL, S. *Non-parametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill, 1956.

SOKOLOV, E. N. *Perception and the conditioned reflex.* Oxford: Pergamon Press, 1963.

WALTERS, R. H., AND DEMKOW, LILLIAN. Timing of punishment as a determinant of response inhibition. *Child Development*, 1963, **34**, 207–214.

WALTERS, R. H., LEAT, MARION, AND MEZEI, L., Inhibition and disinhibition of responses through empathetic learning. *Canadian Journal of Psychology*, 1963, **17**, 235–243.

WALTERS, R. H., AND PARKE, R. D. Influence of response consequences to a social model on resistance to deviation. *Journal of Experimental Child Psychology*, 1964, **1**, 269–280.

WALTERS, R. H., AND PARKE, R. D. The influence of punishment and related disciplinary techniques on the social behavior of children: theory and empirical findings. In B. A. Maher (Ed.), *Progress in Experimental Personality Research.* Vol. 3, New York: Academic Press, 1967, Pp. 179–227.

WALTERS, R. H., PARKE, R. D., AND CANE, VALERIE A. Timing of punishment and the observation of consequences to others as determinants of response inhibition. *Journal of Experimental Child Psychology*, 1965, **2**, 10–30.

WINER, B. J., Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

YOUNG, R. K., AND VELDMAN, D. J. Heterogeneity and skewness in analysis variance. *Perceptual and Motor Skills*, 1963, **16**, 588.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 245–254 (1969)

Psychophysical Scaling of Olfactory Response to the Aliphatic Alcohols in Human Neonates[1]

CAROLYN KENT ROVEE[2]
Brown University

A power function described the relation between successive concentrations of five aliphatic alcohols and response magnitude as measured by integrated stabilimeter activity. Total response level and average threshold concentration are inverse functions of chain length. Infant slopes are higher than those of adults for similar alcohols. These findings are consistent with nonprimate electrophysiological recordings and with human adult scaling of olfactory response.

The limited number of studies describing the olfactory discriminative capacity of human infants probably reflects two factors: methodological and population difficulties, and widely held beliefs that the sense of smell is not mature enough to warrant analysis. Morgan and King (1966, p. 59) note, for example, that at birth taste and smell ". . . apparently are functional, but only on a rudimentary level." Recent studies of infant olfaction have dealt with developmental threshold changes (Lipsitt, Engen, and Kaye, 1963), differential response decrement and recovery to single stimuli (Engen, Lipsitt, and Kaye, 1963), and differential adaptation to components of mixtures (Engen and Lipsitt, 1965). In addition, Engen (1965) found that infants could order a homologous series of alcohols systematically with respect to chain length, both at 100% and threshold concentrations. His response measure was respiratory change versus no-change, precluding determination of differential response magnitudes for solutions between 100% and threshold concentrations.

The present paper describes the use of summated stabilimeter activity,

[1] This research was done in partial fulfillment of the doctoral requirements at Brown University, 1966, while the author was a United States Public Health Service Trainee in Experimental Child Psychology under program no. PHS-MH-5T1-6951. The research was supported by grants (GB-4151, GB-724) from the National Science Foundation to Dr. Trygg Engen, under whose direction the thesis was prepared. The testing laboratory was supported by United States Public Health Service grant no. NB-04268 to Dr. Lewis P. Lipsitt. The author expresses appreciation to Dr. Engen, Dr. Lipsitt, and Dr. Herbert Kaye for their advice and critical comment.

[2] Now at the Department of Psychology, Trenton State College, Trenton, New Jersey 08625.

calibrated for infant weight, as a simple direct measure of response magnitude across concentrations of selected aliphatic alcohols. This measure correlates highly with respiratory change (Levin and Kaye, 1964) and permits valid between-*Ss* comparisons (Rovee and Levin, 1966). The first experiment described below scales newborn response in terms of summated stabilimeter activity across selected concentrations for each alcohol, irrespective of threshold considerations. The second study describes determination of absolute threshold values for the stabilimeter measure.

METHOD

Subjects. The *Ss* were 125 healthy and apparently normal infants from the clinic and private populations of the Providence Lying-In Hospital, tested as they became available with respect to hospital routine. No infant less than 30 hours of age was tested, as olfactory thresholds are still declining prior to this age (Lipsitt *et al.*, 1963). Average age was about 62 hours with a range from 30–107 hours and a standard deviation of about 23 hours. Ten additional *Ss* were discarded from the sample for failing to meet "state" criteria for stimulus presentations. Each *S* was used in only one study and had not participated in previous olfactory tests.

Apparatus and materials. General body movement within the stabilimeter crib (Lipsitt and DeLucia, 1960; Engen, Lipsitt, and Kaye, 1963) was continuously recorded and integrated on separate channels of a Grass S5 Polygraph run at 5 mm/second. On a third channel, respiration was recorded through a pressure transducer-linked Phipps and Bird pneumograph strapped abdomenally; and on a fourth channel, stimulus presentations were indicated by a manually activated event marker (Fig. 1).

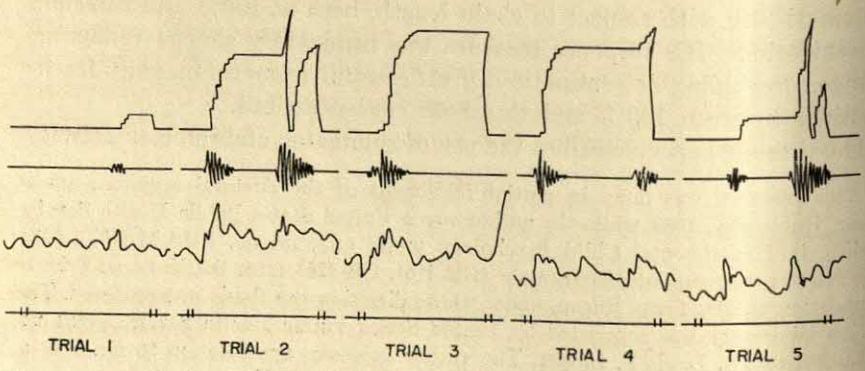


FIG. 1. Sample record of scaling procedure showing manual event marker, respiration, stabilimeter oscillations, and integrated stabilimeter activity. Trials 1–5 represent successive presentation of 6.25, 12.5, 25, 100, and 50% propanol. The intertrial intervals are not shown.

The odorants were propanol, pentanol, hexanol, octanol, and decanol, selected from the linear aliphatic series. The alcohols were supplied by Eastman Organic Chemicals and were redistilled; the diluent was diethyl phthalate, supplied by Fritzsche Bros. A geometric ($\log_2$) dilution series was prepared for five concentrations of each alcohol in 1-cc quantities and maintained in 10 × 75-mm pyrex test tubes, sealed with foil-wrapped corks. The cotton-swathed end of a glass rod anchored to the cork remained partially suspended in the odorant when the cork was in place. Stimuli were presented by placing the saturated cotton tip in position approximately 5 mm beneath the *S*'s nostrils. At no time was the cotton tip in contact with *S*'s facial region, and slight head movements which occurred infrequently within the stimulation interval did not disrupt normal stimulation. Although a single blind procedure may have been preferable, the reliability and validity of the present method has been well documented (Lipsitt *et al.*, 1963; Engen *et al.*, 1963; Engen, 1965).

Procedure. All the tests were conducted in a semi-darkened, quiet, ventilated laboratory in the Lying-In Hospital between 10:15 AM and 11:15 AM, after the regular 9:30 AM to 9:50 AM feedings. An assistant carried each infant from the nursery to the testing room, placed the *S* supine in the stabilimeter crib, and fastened the pneumograph. The *S* was then tightly swaddled to increase the probability of maintaining a quiescent state during testing (Richmond, Lipton, and Steinschneider, 1962). The stabilimeter was then calibrated for the *S*'s weight.

All trials consisted of 10-second stimulus presentations with intertrial intervals of approximately 60 seconds. Trial initiation was conditional upon a 10-second prestimulation period of regular respiration, minimal stabilimetric activity, and closed eyes. The first trial was usually begun within 5 minutes of the initial preparatory activities. If *S* developed respiratory irregularities (e.g., hiccoughs, crying, jerky breathing), testing was discontinued. A response was defined as stabilimetric displacement within the 10-second stimulation interval and the total cumulative millimeters of displacement constituted the magnitude of response.

RESULTS

Experiment 1

Each subgroup of 15 *S*s was tested on the dilution series of one of the five alcohols. Every *S* received all five concentrations in random order, the only restriction being that each stimulus must occur first equally often.

All psychophysical functions were obtained by the method of least squares, using geometric means. To permit logarithmic transformation, all responses of zero magnitude were assigned the value of 0.3 mm, the

smallest detectable increment of the ink writer. A discussion of the use of 0.3 mm is contained elsewhere (Engen, Cain, and Rovee, 1968).

As summarized in Fig. 2, the relationship between response magnitude (log-mm response) and stimulus intensity (log percent concentration) is approximately linear, or a power function, for each alcohol tested. These data are consistent with audition data (Bartoshuk, 1964) and adult olfaction scaling data (Engen, 1965; Engen *et al.*, 1968). Further, total response level appears to be an inverse function of the number of carbon atoms in the alcohol chain, corresponding to the findings of Engen (1965) for respiratory measures.

Coefficients of determination (r^2) indicate that the goodness of fit is best over concentrations of the short chain alcohols, with the exception

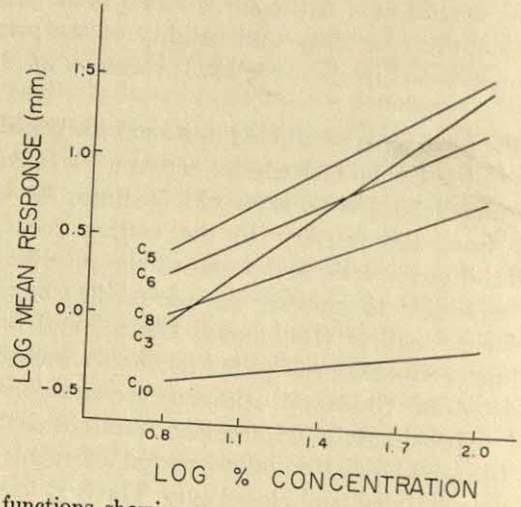


FIG. 2. Power functions showing response magnitude as a function of number of carbons and stimulus intensity, Experiment 1. The odorants were propanol (C_3), pentanol (C_5), hexanol (C_6), octanol (C_8), and decanol (C_{10}).

of pentanol, and deteriorates as chain length increases. However, as the functions get progressively flatter with increasing chain length, the r^2 value must get smaller, as it indicates only relative variability between stimulus and response, and there is less variation in the height of response across concentrations of longer chain length alcohols. To elaborate, as slopes decrease, variability "takes over," or becomes larger relative to stimulus concentration. Inspection of the averaged data within each series revealed no response overlap across successive alcohols with the exception of a propanol-pentanol reversal at all concentrations but 100% and a pentanol-hexanol reversal at 6.25%. In addition, the standard error does not exceed a fourth of a log unit within any concentration.

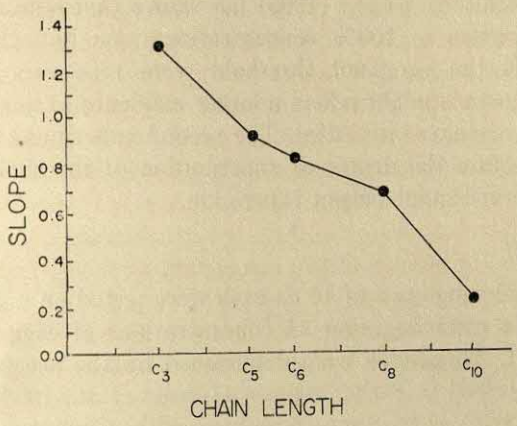


FIG. 3. Power function slope as a function of chain length.

Since magnitude of response varies directly with stimulus intensity and inversely with chain length, the input-output relations might be compared across alcohols by the slope of each function (Table 1). As Fig. 3 shows, the exponent or slope decreases monotonically with alcohol chain length. Engen (1965) found a similar relation between newborn respiratory response to 100% concentrations and chain length. However, the slope values reported here are much higher than those obtained by Engen from adults using the method of magnitude estimation for pentanol (0.42) and octanol (0.13) (Engen *et al.*, 1968). Stevens (1961) has also reported low exponents for olfactory stimuli. The slopes, however, appear to differentiate the alcohols numerically only. There is considerable overlap (Fig. 2) between slope points at various concentration levels of different odors. The propanol series, with the largest exponent or steepest slope, is indistinguishable from pentanol at higher concentrations, from hexanol at intermediate concentrations, and from octanol at

TABLE 1

COMPARISON OF ORIGINAL AND CORRECTED SLOPES (METHOD OF LEAST SQUARES), COEFFICIENTS OF DETERMINATION, AND THRESHOLD VALUES

Number of carbons	Experiment 1		Experiment 2			
	Slope	r^2	Corrected slopes	r^2	Mean RL	Standard error
Propanol-3	1.303	.96	0.602	.75	16.25%	1.90
Pentanol-5	0.936	.78	0.459	.77	7.50%	1.16
Hexanol-6	0.845	.89	0.561	.86	4.37%	1.04
Octanol-8	0.704	.83	0.640	.80	1.40%	0.30
Decanol-10	0.233	.56	0.242	.56	0.58%	0.13

lower concentrations. Engen (1965) has shown that stimuli eliciting more vigorous responses to 100% concentrations also have higher detection thresholds. If the propanol threshold were relatively high, averaged response magnitude might reflect a larger incidence of nonresponse among *Ss* to lesser stimulus intensities. The second experiment was conducted, then, to ascertain the degree of contribution of threshold effects to the initially observed input-output expression.

Experiment 2

Five alcohol subgroups of 10 *Ss* each were tested on a geometric (log_2) dilution series extending over 11 concentrations of each alcohol used in Experiment 1. Thresholds were determined by the ascending method of limits as described in earlier reports (Lipsitt *et al.*, 1963; Engen, 1965). On the first trial, each *S* was presented with a member of the dilution series which was known to be below the threshold concentration for that alcohol. On subsequent trials, concentrations within the series were successively doubled. Threshold for a given *S* was defined as that concentration to which the infant first responded. One concentration higher than that evoking the first response was always presented, and in every instance also evoked a response (Fig. 4).

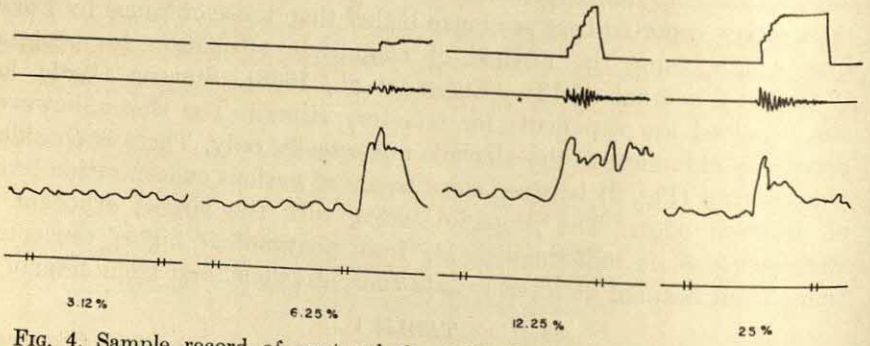


FIG. 4. Sample record of pentanol threshold determination in ascending series. Threshold for this *S* was defined as 12.5% pentanol; however, note the late respiratory response to 6.25%.

Average threshold concentration was found to be a decreasing function of the number of carbons in the alcohol chain (Fig. 5), as reported by Engen (1965), although the values for corresponding stimulus points are higher over all concentrations in the present study. This suggests that more intense stimulation is required to produce movement as opposed to respiratory disruption. Inspection of individual records indicated that the majority of infants did exhibit breathing changes prior to the movement threshold concentration for that *S*. It is also possible that the

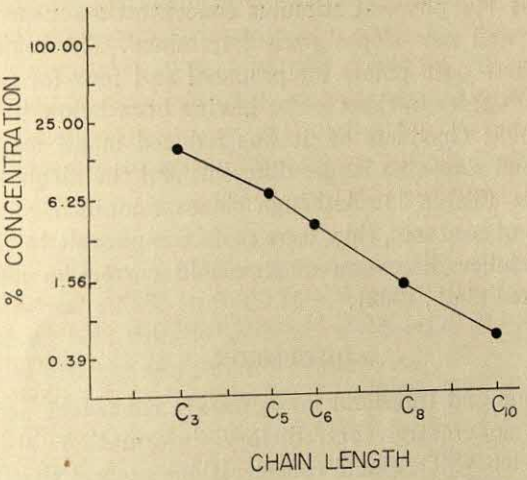


FIG. 5. Stimulus threshold intensity (percentage concentration) as a function of chain length.

swaddling used to maintain state suppressed or obscured minimal response. Response variability varied directly with threshold intensity and inversely with chain length.

In order to describe the obtained response magnitude across only the effective or suprathreshold range of concentrations, each concentration was respecified in terms of its effective stimulus intensity according to

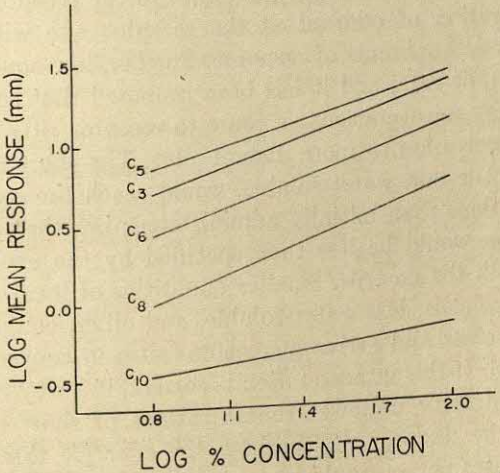


FIG. 6. Response magnitude as a function of the effective supra-threshold stimulus intensity over five homologous alcohols. These are threshold-corrected functions from Fig. 2.

the formula log (% physical stimulus concentration–mean % threshold concentration) and new slopes were determined. The revised functions include only three data points for propanol and four for pentanol (Fig. 6), the omitted values in these series having been below the average detection thresholds. Goodness of fit was reduced in all instances that it was affected, and slopes no longer differentiated the alcohols in any systematic fashion (Table 1). Although slopes remained relatively steep, especially that of pentanol, they were more comparable to those reported for adults. A fuller discussion of threshold correction effects appears elsewhere (Engen *et al.*, 1968).

DISCUSSION

Although slope and threshold measures are consistent across alcohols, they provide apparently contradictory information about stimulus strength or efficiency. Threshold considerations suggest that the strongest stimulus is the one which is effective at the lowest concentration; however, the stimulus eliciting a response of the greatest magnitude is not lowest in threshold concentration. Paradoxically, propanol produces the most vigorous response and also has the highest threshold of the alcohols studied; decanol can produce only low-magnitude responses but is effective at very low concentrations.

Electrophysiological studies on animals (Tucker, 1963; Ottoson, 1956, 1958) using the same homologous series have shown a similar relation between response magnitude and chain length. Tucker (1963) has indicated that both delivery rate of molecules to the receptors and the effective concentration of odorant at the receptor site within a critical time are crucial determinants of response. Further, low aqueous solubility correlates with flow rate, and it has been proposed that some odorant is lost in the watery nasal mucous en route to receptor sites, more odorant being lost in progression to more distant sites. The short chain alcohols, being more volatile and water-soluble, would reach the receptor sites in lesser concentrations than initially administered; i.e., the *effective* stimulus concentration would be less than specified by the experimenter and presented beneath the nostrils. Smaller quantities of long chain alcohols, which are less volatile, less water-soluble, and oilier, could travel faster to stimulation points and arrive at receptor sites in concentrations relatively unchanged. If this interpretation is correct, if a way could be found to ensure the integrity of lower concentrations of short chain alcohols on the way to the receptor site, perhaps the *effective* threshold concentrations at the sites per se would be lower than presently thought.

A second, related explanation of the paradox involves the trigeminal nerve, thought to be responsible for short chain stimulation effects. Either

the properties of trigeminal receptors may be such as to require concentrations higher than those which can stimulate pure olfactory receptors, or trigeminal sites may be less accessible. Tucker (1963) has indicated that trigeminal receptors are less exposed to the gaseous environment in the nasal cavity than are the olfactory receptors, and has also suggested that trigeminal stimulation produces reflexive changes in the nose which reduce accessibility to olfactory receptors and increase mucosal secretion and flow. He has suggested that trigeminal response to threshold concentrations "probably appears first for smaller members of the aliphatic series" (1963, p. 66). It is likely, then, that the apparent paradox observed behaviorally is simply a reflection of differential interactions between odorant and receptor apparatus.

The present findings are notable for the consistencies they demonstrate between electrophysiological data on subhumans, behavioral data on nonverbal human newborn, and scaling data from verbal reports of adult humans. Further, they show that the human newborn has a much keener olfactory apparatus capable of making finer discriminations than previously thought. The unique position of the infant in terms of past olfactory experience and discriminative ability suggests their future use as Ss in investigations of basic olfactory phenomena.

REFERENCES

Bartoshuk, A. K. Human neonatal cardiac responses to sound: a power function. *Psychonomid Science*, 1964, **1**, 151–152.

Engen, T. Psychophysical analysis of the odor intensity of homologous alcohols. *Journal of Experimental Psychology*, 1965, **70**, 611–616.

Engen, T., Cain, W. S., and Rovee, C. K. Direct scaling of olfaction in the newborn infant and the adult human observer. In N. Tanyolaç (Ed.), *Theories of Odor and Odor Measurement*. Istanbul, Turkey: Robert College, Research Center, 1968. Pp. 271–296.

Engen, T., and Lipsitt, L. P. Decrement and recovery of responses to olfactory stimuli in the human neonate. *J. comp. physiol. Psychol.*, 1965, **59**, 312–316.

Engen, T., Lipsitt, L. P., and Kaye, H. Olfactory responses and adaptation in the human neonate. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 73–77.

Levin, G. R., and Kaye, H. Non-nutritive sucking by human neonates. *Child Development*, 1964, **35**, 749–758.

Lipsitt, L. P., and DeLucia, C. A. An apparatus for the measurement of specific response and general activity of the human neonate. *American Journal of Psychology*, 1960, **73**, 630–632.

Lipsitt, L. P., Engen, T., and Kaye, H. Developmental changes in the olfactory threshold of the neonate. *Child Development*, 1963, **34**, 371–376.

Morgan, C. T., and King, R. A. *Introduction to Psychology*. New York: McGraw-Hill, 1966.

Ottoson, D. Analysis of the electrical activity of the olfactory epithelium. *Acta Physiologica Scandinavia*, 1956, **35**, (Suppl. 122), 1–83.

OTTOSON, D. Studies on the relationship between olfactory stimulating effectiveness and physico-chemical properties of odorous compounds. *Acta Physiologica Scandinavia*, 1958, 43, 167–181.

RICHMOND, J. B., LIPTON, E. L., AND STEINSCHNEIDER, A. Observations on differences in autonomic nervous system function between and within individuals during early infancy. *Journal of Child Psychiatry*, 1962, 1, 82–91.

ROVEE, C. K., AND LEVIN, G. R. Oral "pacification" and arousal in the human newborn. *Journal of Experimental Child Psychology*, 1966, 3, 1–17.

STEVENS, S. S. To honor Fechner and repeal his law. *Science*, 1961, 133, 80–86.

TUCKER, D. Olfactory, vomeronasal and trigeminal receptor responses to odorants. In Y. Zotterman (Ed.), *Olfaction and Taste*. Oxford: Pergamon Press, 1963. Pp. 45–69.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 255–264 (1969)

A Developmental Study of Sequence Structure in Binary Prediction[1]

NANCY A. MYERS AND JEROME L. MYERS
University of Massachusetts

Sixteen groups of 20 *Ss* each were run individually in a binary prediction situation with each alternative occurring equally often. Four age groups were employed: *Ss* were either in kindergarten, second, fourth, or sixth grades. For half the *Ss* at each age level the events occurred in runs of either 1 or 5; the other *Ss* observed events which occurred in runs of either 4 or 5. The probability of occurrence of the long run (β) was also manipulated: for half the *Ss* receiving each set of run lengths there were equal numbers of short and long runs ($\beta = .5$); for the other *Ss* $\beta = .2$. The results indicated that all groups showed some sensitivity to run length characteristics. However, errors of all types decreased as age increased, and 4–5 sequences were followed more accurately than 1–5 sequences. The results were discussed in terms of their implications for a general model of binary choice behavior.

The binary prediction paradigm requires the *S* to predict which of two events will occur on each of a series of trials. Recent developments in the binary prediction literature have been in the direction away from noncontingent probabilistic sequences and toward the study of more carefully controlled sequential structures. With adults, a wide variety of both partially and completely learnable sequences have been employed in attempts to isolate the effective stimuli and response tendencies in binary prediction (e.g., Gambino and Myers, 1967; Restle, 1966). With children, too, there is increasing evidence of some sensitivity to at least some characteristics of sequences. Thus, Craig and Myers' (1963) kindergarten *Ss* alternated more with a 60:40 than with an 80:20 sequence, and Bogartz' (1966) 4- and 5-year-olds followed single alternations quite well, at least in early trials, and were obviously affected by changes in sequential structure. Jones (1967) found that both nursery and kindergarten *Ss*' responses reflected the pattern of highly discriminable event sequences. She exposed *Ss* to sequences with event repetition probabilities

[1]This research was supported by funds from a University of Massachusetts Faculty Research Grant and NIMH Grant MH-03803-06 and NSF Grant GS-386. The authors thank the Research Computing Center, University of Massachusetts for assistance in analyzing data.

of either .1 or .9 for 100 trials and for the second 100 trials transferred half the *Ss* to the opposite event repetition probability, while continuing the others at the same repetition probability level. When the repetition probability level changed, responses showed only limited dependency upon the new event sequence. However, kindergarten *Ss* followed any sequence better than nursery *Ss* and both age groups followed predominately repeating sequences more accurately than predominately alternating sequences.

The present study was designed to explore further those characteristics of event sequences to which children may respond. In particular, we were interested in sensitivity to run length, because Myers, Butler, and Olson (in press) had observed such sensitivity in adults. Following the Myers *et al.* procedure, the events always appeared in runs of two standard lengths, either 4 and 5 or 1 and 5. For some *Ss* the two run lengths occurred equally often; for others, the shorter run length occurred more frequently. In addition, four age levels were employed, to permit evaluation of developmental aspects of performance.

METHOD

Apparatus. The event box consisted of two 1½ v. red event lights mounted in parallel on the front of an 8- × 12- × 2-inch grey box, one 3 inches from the left, the other 3 inches from the right edge. Below each light was a 3½- × ½-inch slot through which one of 200 3- × 2-inch playing cards, blank on both sides, could be placed to make a response. A grey panel attached to the back of the event box concealed the control unit from *S*. The control unit contained a buzzer that was used to indicate when choices should be made. One switch on the control unit sounded the buzzer and two other switches turned on the event lights.

Design. *Ss* were randomly assigned to one of sixteen experimental groups differing with respect to age and two characteristics of the event sequences employed, in a $4 \times 2 \times 2$ factorial design. There were 20 *Ss* in each group, half males and half females. Four age groups were run: *Ss* were either in kindergarten, second, fourth, or sixth grades. For half the *Ss* at each age level the events occurred in runs of either one or five; the other *Ss* observed events which occurred in runs of either four or five. The other aspect of event sequences which was manipulated was β, the probability of occurrence of the *long* run. For half the *Ss* receiving each set of run lengths, β was .5, i.e., there were equal numbers of short and long runs. For the other *Ss*, β was .2, and thus short runs predominated.

Event sequences were generated in two blocks of approximately 100 trials each, for each of four run length–β combinations, with the one additional requirement that each light occur equally often.

Procedure. The *S*s were tested one at a time in what they were told was a "guessing game." After *E* signaled the beginning of a trial with the buzzer, *S* predicted which one of the two lights would come on by taking a card from the stack of 200 in front of him and placing it in the slot under the appropriate light. For the first five trials, *S* was required to point to the light and then put the card in the slot. This procedure was used in order to ensure that *S* understood the instructions. The game was played at *S*'s own rate, and *S* was required to look at the lights before *E* ended the trial. It was emphasized in the instructions that the order of the lights was predetermined and that the card did not turn on the lights. Each session lasted approximately 30 minutes.

Subjects. The *S*s were 320 children attending public school classes in Northampton, Massachusetts.[2] Eighty children were run at each of four grade levels: kindergarten, second, fourth, and sixth grades. No child who had repeated a grade was employed as *S*. The mean age, and the age range, in years and months was respectively: kindergarten, 5–7, 4–10 to 6–1; second grade, 7–8, 6–8 to 8–3; fourth grade, 9–6, 8–9 to 10–1; sixth grade, 11–7, 10–9 to 12–3.

RESULTS

Three measures will be considered in addition to the usual sequential analysis of repetition responses as a function of the preceding run of events. Anticipatory errors are failures to predict the preceding event when the probability of an event repetition is one, and perserverative errors are failures to predict the alternative event when the probability of an event switch is one. The proportion of repetition responses at the uncertainty point represent repetitions after runs of length one in 1–5 groups and repetitions after runs of length four in 4–5 groups. These measures are of particular theoretical import because of their sensitivity to sequence structure (Gambino and Myers, 1967) and because previous models which viewed runs as the critical stimuli were unable to predict the occurrence of such errors (Restle, 1966).

Anticipatory errors. An anticipatory error is defined as the prediction of the alternative event when the probability of a run continuing is one. Such errors can occur following runs of length 2, 3, and 4, in 1–5 groups and following runs of length 1, 2, and 3, in 4–5 groups. The proportion of anticipatory errors made by each of the experimental groups at each grade level may be seen in Fig. 1, and the results of an analysis of vari-

[2] The authors thank John Buteau, Superintendent of Schools, and Robert Moriarty, Elementary School Supervisor, Northampton, Mass., for their cooperation in providing facilities and *S*s for this study, and Mrs. Florence Mador for assistance in running *S*s.

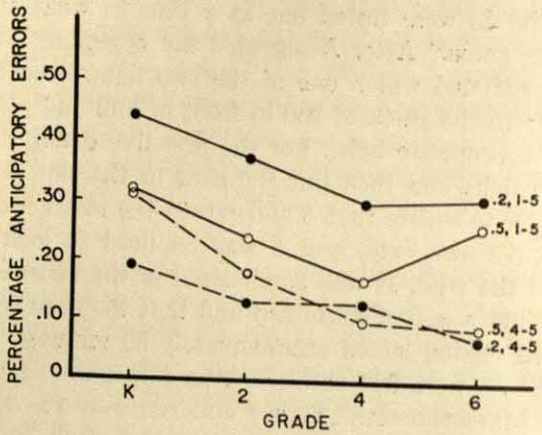


FIG. 1. Percentage anticipatory errors as a function of grade, for each β-RL combination.

TABLE 1

SUMMARY OF ANALYSES OF VARIANCE OF THREE DIFFERENT MEASURES

Source	*df*	Anticipatory errors		Perseverative errors		Repetitions at uncertainty point	
		MS	*F*	MS	*F*	MS	*F*
β (Proportion of long runs)	1	.1918	5.7[a]	1.3995	15.5[c]	6.3090	114.5[c]
G (Grade)	3	.6667	19.8[c]	1.3331	14.7[c]	.4880	8.8[b]
R (Run length)	1	3.7130	110.1[c]	4.4051	48.7[c]	.1545	2.8
T (Trial block)	1	1.1391	199.8[c]	.0328	1.0	.2125	11.2[c]
βG	3	.0513	1.5	.1877	2.1	.4690	8.5[b]
βR	1	.9048	26.9[c]	.4456	4.9[a]	.1445	2.6
GR	3	.0523	1.6	1.0750	11.9[c]	.0772	1.4
βT	1	.0214	3.8	.0023	<1	.1686	8.9[b]
GT	3	.0224	3.9	.0046	<1	.0014	<1
RT	1	.0017	<1	.0767	2.4	.0004	<1
βGR	3	.0581	1.7	.5768	6.4[a]	.2492	4.5[a]
βGT	3	.0081	1.4	.0054	<1	.0364	1.9
βRT	1	.0013	<1	.0722	2.3	.0688	3.6
GRT	3	.0136	2.4	.0545	1.7	.0332	1.7
S/βGR	304	.0337	—	.0904	—	.0551	—
βGRT	3	.0058	1.0	.0045	<1	.0163	<1
S × T/βGR	304	.0057	—	.0320	—	.0190	—

[a] Significant at the .05 level.
[b] Significant at the .01 level.
[c] Significant at the .001 level.

ance performed on these data may be found in Table 1.[3] Many more anticipatory errors were made by the 1–5 run length (RL) groups than by the 4–5 RL groups, F (1,304) = 110.1, $p < .001$. There are also significantly more anticipatory errors in the .2 than in the .5 groups, F (1,304) = 5.7, $p < .05$. This effect is due primarily to the 1–5 groups, and is actually reversed in the younger 4–5 groups; this $\beta \times$ RL interaction is significant at the .001 level, F (1,304) = 26.9. Anticipatory errors also decreased with age [F (3,304) = 19.8, $p < .001$], and with trial block [F (1,304) = 199.8; $p < .001$].

Perseverative errors. A perseverative error is defined as the prediction of the preceding event (or the failure to predict the alternative event) when the probability of a run continuing is zero, i.e., after a run of five events. The proportion of perseverative errors made by each of the experimental groups at each grade level may be seen in Fig. 2, and the

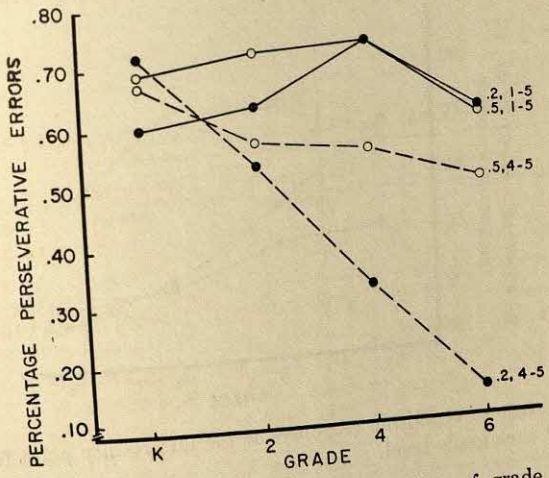


FIG. 2. Percentage perseverative errors as a function of grade for each β–RL combination.

summary of the appropriate analysis in Table 1. The 1–5 RL groups made many more perseverative errors than the 4–5 groups, F (1,304) = 48.7, $p < .001$. It is interesting to note that, with the total number of trials equated at approximately 200, 1–5 groups observed considerably more runs of length 5 than did 4–5 groups, and therefore, if practice was a factor, should have shown less, not more perseverative errors.

[3] All error proportions reported in the Results were computed by dividing the number of such errors by the number of times such an error could have occurred.

Averaging over all grade and RL conditions, perseverative errors decreased with a decrease in the proportion of long runs from .5 to .2; F (1,304) = 15.5, $p < .001$. A significant $\beta \times$ RL interaction, F (1,304) = 4.9, $p < .05$, supports the interpretation that the β effect is due primarily to the large decrease in perseverative errors for the .2, 4–5 groups compared to the .5, 4–5 groups. Perseverative errors decreased significantly with age, F (3,304) = 14.7, $p < .001$. The decrease with age is primarily in the 4–5 groups, particularly the .2, 4–5 groups; the RL $\times$ Grade and RL $\times$ Grade $\times \beta$ interactions are both significant; F (3,304) = 11.9, $p < .001$, and F (3,304) = 6.4, $p < .05$, respectively.

Repetitions at the uncertainty point. The proportion of repetition responses at the uncertainty point (i.e., following runs of length one in the 1–5 groups and of length four in the 4–5 groups) was also analyzed. The appropriate analysis is summarized in Table 1 and the relevant results may be seen in Fig. 3 which shows the percentage of repetitions at the

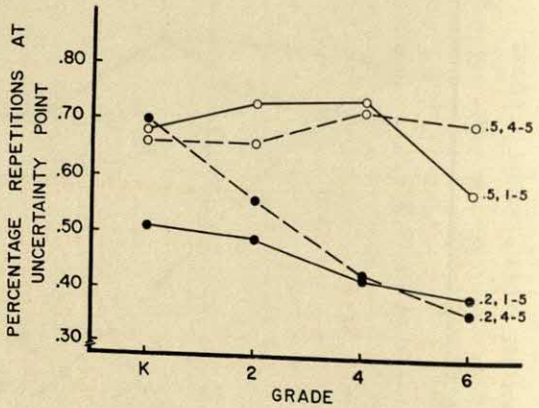


FIG. 3. Percentage of repetition responses at the uncertainty point for each experimental group at each grade level.

uncertainty point for each experimental group at each grade level. As the proportion of long run lengths decreased, the proportion of repetitions at the choice point decreased, F (1,304) = 114.5, $p < .001$. There was an overall decrease in uncertainty point repetitions with age, F (3,304) = 8.8, $p < .01$. This was due mainly to a decrease with age in repetitions among the .2 groups, and particularly in the .2, 4–5 RL groups; the $\beta \times$ Grade and $\beta \times$ Grade $\times$ RL interactions were both significant, F (3,304) = 8.5, $p < .01$, and F (3,304) = 4.5, $p < .05$. Not visible in the figure is a slight increase in uncertainty point repetitions in the second block of trials, F (1,304) = 11.2, $p < .001$. This effect was almost entirely due to the increase in repetitions shown by the .5

groups in block 2, while the .2 groups made almost the same proportion of repetition responses at the uncertainty point in the two trial blocks; the $\beta \times$ Trial Block interaction was significant at the .01 level, $F(1,304) = 8.9$. At all ages, the proportion of repetitions was always greater than the values of β, indicating a tendency to predict the repetition of an event more often than it happens.

Run curves. The run curves in Fig. 4 show the percentage of repetition responses as a function of the length of the preceding run of events for each grade level, for each of the four β–RL combinations. The measures

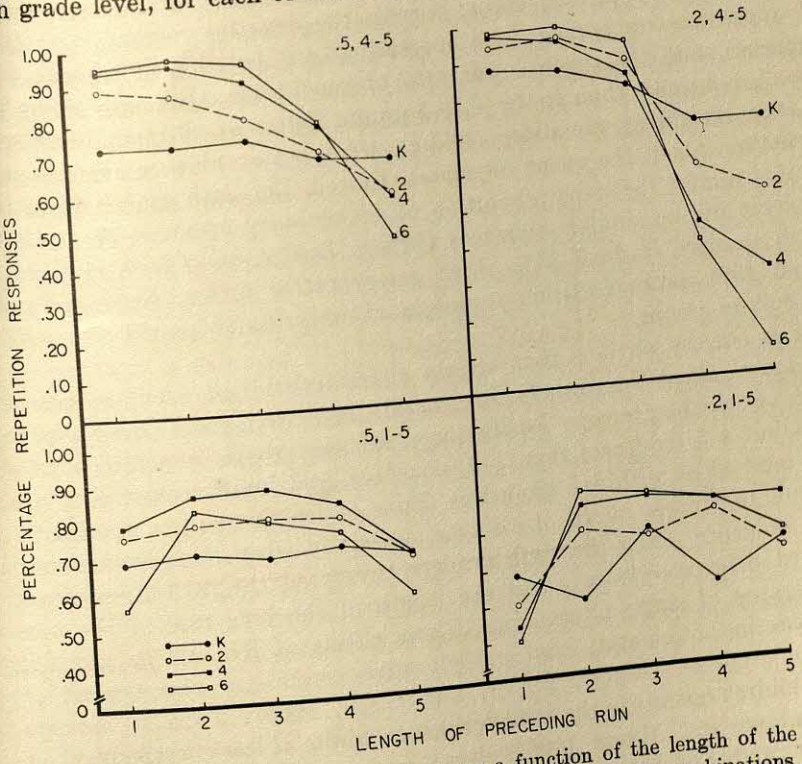


FIG. 4. Percentage of repetition responses as a function of the length of the preceding run of events, for each grade level, for each of the β–RL combinations.

analyzed earlier partially determine these curves, but Fig. 4 serves as a quick visual summary of the differences in sensitivity of the four age groups to the regularities of the sequences, and their differential sensitivity to different sequences. In both top panels, all grade levels show a decline in repetition responses following runs of length four, and a further decline following runs of length five. This finding becomes more marked with age, and it is even more obvious in the .2, 4–5 RL con-

dition than in the .5, 4–5 RL condition. The 1–5 RL conditions shown in the bottom half of Fig. 4 reveal much less sensitivity to these sequences at all grade levels, and this is particularly true in the .2, 1–5 RL condition, where runs are of length one 80% of the time.

DISCUSSION

The results of this experiment demonstrate that children are sensitive to more than just the last event in the binary prediction situation. Even the youngest age group was clearly picking up some information about run lengths in some sequences; witness the negative recency (decrease in repetition responding with increasing run length) in the .2, 4–5 sequence, and the much lower rate of repetitions after one event in the .2, 1–5 group than in the .5, 1–5 group. Errors of all types decreased as age increased; the older children were better able to process information about the event sequences. Indeed, the sixth graders demonstrated almost the same proportion of anticipatory errors as did college students run on similar sequences (Myers, Butler, and Olson, in press). Sixth graders made 7–33% more perseverative errors, however, and about 15% more repetition responses at uncertainty points, than did the college groups.

It is equally obvious that certain characteristics of event sequences are more salient than others. In general, there were more perseverative than anticipatory errors. Furthermore, all age groups were better able to follow 4–5 sequences than 1–5 sequences, and improvement with age was most clear with 4–5 sequences. Runs of length one are apparently difficult to process for *Ss* of all ages, especially when they predominate in a sequence. Even the sixth graders, presented the .2, 1–5 sequence, showed little knowledge about the length of the long run, or the preponderance of single alternations. These effects of RL and β, and differences in anticipatory and perseverative errors are also typical of college students (Myers, Butler, and Olsen, 1969). Thus it appears that children operate in the same fashion as adults, if less effectively.

Gambino and Myers (1967) have recently proposed a model for binary prediction behavior which predicts many of the effects observed above. They assume that runs of events are discriminative stimuli but that *Ss* cannot discriminate perfectly. Because of this imperfect perception, reinforcement of the prediction that a run length will continue or break off may generalize and serve as reinforcement for the prediction of other run lengths. The degree of generalization is determined by the absolute difference between the reinforced run length and the particular other run length of concern. This view can account for the

greater number of anticipatory errors made by the 1–5 groups. There is one basic structural difference between the two sequences: In 1–5 sequences the anticipation error points are sandwiched between two break-off points (i.e., they can occur following runs of length 2, 3, or 4), while in 4–5 sequences runs of length 5 are at least two points removed from the closest anticipatory error point (anticipatory errors can occur following runs of length 1, 2, or 3). Therefore, according to the generalization model, there should be more switching at anticipatory points in 1–5 groups than in 4–5 groups. The concept of a generalization gradient over run lengths would also predict more perseverative errors in the 1–5 groups; the breaking off of a run of length one yields less generalized switching at runs of length five than does the breaking off of a run of length four. The fact that differences in the error rates of 1–5 and 4–5 groups are more marked at $\beta = .2$ is consistent with this account.

The results would be equally well accounted for by the assumption that the primary source of confusion is miscounting, that is, that the subject misperceives the current run length and that only the probability associated with a run of the perceived length is influenced by the trial outcome. If it is assumed that the subject's count is most likely to be off by a small amount, miscounting would produce results very similar to those that were attributed to generalization. Since it is also highly probable that counting accuracy improves with age, this interpretation would account equally well for the developmental changes.

In summary, the ability to process structured sequential information improves continuously with age, and shows the same relative effects of experimental manipulation at all age levels, including adulthood. A model which views event runs as imperfectly discriminated stimuli, and attributes age level differences to differences in capacity to discriminate among run lengths, provides a reasonable first account of the data.

REFERENCES

Bogartz, R. S. Variables influencing alternation prediction by preschool children: I. Previous recurrent, dependent, and repetitive sequences. *Journal of Experimental Child Psychology*, 1966, 3, 40–56.

Bogartz, R. S., and Pederson, D. R. Variables influencing alternation prediction by preschool children: II. Redundant cue value and intertrial interval duration. *Journal of Experimental Child Psychology*, 1966, 4, 211–216.

Craig, G. J., and Myers, J. L. A developmental study of sequential two-choice decision making. *Child Development*, 1963, 34, 483–494.

Gambino, B., and Myers, J. L. The role of event runs in probability learning. *Psychological Review*, 1967, 74, 410–419.

JONES, S. The effect of repetition probabilities on children's two-choice learning. Unpublished doctoral dissertation, University of Massachusetts, July, 1967.

MYERS, J. L., BUTLER, P., AND OLSON, D. Run lengths and probabilities in binary prediction. *Journal of Mathematical Psychology*, 1969, **6**, in press.

RESTLE, F. Run structure and probability learning: disproof of Restle's model. *Journal of Experimental Psychology*, 1966, **72**, 382–389.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 265–273 (1969)

Number and Color Responses in the Young Child[1]

ROSE GINSBERG
San Jose State College

Five-year-old kindergarten children were taken through a successive discrimination task to which they could respond correctly either on the basis of number or color. Subsequently they were shifted to a number-relevant, color-irrelevant; or a color-relevant, number-irrelevant task. The results indicated that *Ss* in the pretransfer trials responded to the number and not to the color dimension.

To ensure that *Ss* were responding to the number dimension rather than possible density or form attributes, a second experiment was performed. In Experiment 2 the same aged children learned the Experiment 1 task, with number-only relevant, and the learners were transferred to a much more difficult task with the same dimension still relevant. A control group learned the more difficult task only. All learners in the experimental group transferred successfully, but no *Ss* in the control group reached criterion.

It was concluded that (a) *Ss* in Experiment 1 were responding to the number dimension and (b) if *Ss* can respond to the number dimension in a very simple task, they will transfer successfully to the same dimension in a task which is too difficult without the preswitch practice.

Recent studies have shown that the five- and six-year-old can, in an experimental situation, learn quite difficult number concepts (Suppes and Ginsberg, 1962, 1963) and in the practical situation some formal teaching of number concepts is now commonly introduced at the kindergarten level (*Sets and Numbers*, 1936; SRA, 1961). Experimental investigations with this age group, on the other hand, have been largely focused on form, color, and sometimes size (Suchman and Trabasso, 1966; Kagan and Lemkin, 1961; Brian and Goodenough, 1929) those dimensions of a stimulus traditionally assumed to have special importance for the preschooler. Lee (1965) included a number concept in an investigation of concepts in young children. Using a modified oddity task and five examples of each concept, Lee reports that with all ages involved (3–6 and 6–5) more errors were made with number than with color or size.

[1]This study was performed when the investigator was at Queens University, Kingston, Ontario, Canada, and was supported by Mental Health Research Grant 605-5-373. The author thanks the Kingston Board of Education for their cooperation in providing subjects from the Kingston Public Schools.

The present experiment makes a direct comparison of the prepotency of number and color in the five-year-old child. Initially the child is presented with a series of stimuli to which he can respond correctly on the basis either of number or color. Subsequently he is transferred to a situation in which either number or color alone is relevant, and his performance on the transfer task is used to identify the dimension to which he was responding on the original learning task.

METHOD

Subjects

The *S*s were 48 kindergarten children (mean age 65 months) from two public schools in Kingston, Ontario, Canada. None of the children had previously been exposed to a formal presentation of number concepts. Within each school, the children were randomly assigned to one of three groups.

Apparatus

The apparatus consisted of a 16½- × 13½-inch wooden screen with two 5- × 4-inch windows approximately at the young child's eye level when he was seated on a low chair. On *S*'s side, 1 inch beneath each window, there was a small white reinforcement light with a response button on the base of the apparatus, 1 inch beneath and in front of each reinforcement light. The stimuli (a picture in each window) were presented on each trial by inserting (on *E*'s side) an 8- × 11-inch card. Reinforcement was controlled by a lever on *E*'s side and two response lights on *E*'s side indicated which response *S* made on each trial. The source of power was a six volt battery.

Design

On each trial the child was presented with two pictures and was required to press one of the response buttons if the pictures "went together" and the other if they did not "go together." Following a correct button press, the reinforcement light immediately above the button went on. In all three groups on Day 1, the pictures "went together" if each picture showed the same number of objects. In Groups 1 and 2, the "equal-number" pictures were also in one color, the "unequal-number" pictures in a different color. On Day 1 therefore, in Groups 1 and 2, both number and color were relevant and the child could respond correctly on the basis of color or number. In Group 3 all pictures were

black on a white background and the child could, presumably, respond correctly only on the basis of number.

On Day 2, the procedure for Group 3 remained unchanged, so that number was still the only relevant dimension. For Group 1 on the other hand, number was relevant and color irrelevant; for Group 2, color was relevant and number irrelevant. Table 1 summarizes the experimental design.

TABLE 1
EXPERIMENTAL DESIGN: EXPERIMENT 1

Groups	Day 1	Ss meeting criteria	Day 2[a]	Ss meeting criteria
1(N = 16)	Number- and color-relevant	11	Number-relevant	9
2(N = 16)	Number- and color-relevant	12	Color-relevant	0
3(N = 16)	Number only relevant	12	As Day 1	12

[a] Only Ss meeting criterion on Day 1 were taken through Day 2 procedure.

Stimuli

Six stamps were used to make 36 stimulus pictures—a triangle, circle, balloon, box, star, and cross with the pictures varied in size from the circle (with a diameter of ½ inch) to the rectangle (½ × 1¼-inches). Each of the two pictures presented on a single trial consisted of from one to three objects. On any one trial the objects presented in the two pictures were the same so there might, for example, be three balls and three balls on one trial, on another trial two stars and one star and so on.

With one to three objects in each picture there were six possible combinations and each combination was represented by all six different objects. Of the "unequal-number" combinations, the picture with less objects was presented on half of the trials in the left-hand window, on the other half in the right-hand window. In deference to Piaget's (1953) suggestion that the younger child judges cardinal equivalence on the basis of the lengths of two rows of objects, the actual length of the "row" of objects presented in each pair of pictures was not consistent—the lesser number of objects in a pair of pictures of different cardinal value sometimes being the longer of the two.

Day 1. For half of the Ss in Group 1 and half of the Ss in Group 2, the objects in the "equal-number" pictures were red; in the "unequal-number" pictures they were blue. The colors were reversed for the remaining Ss.

Day 2. The stimuli, on Day 2, were again identical for both Groups 1 and 2 but on this day half of the "equal-number" and half of the

"unequal-number" pictures were red, half were blue. Thus, using the same stimuli, either number alone or color alone could be reinforced. For Group 1 (number-relevant) "equal-number" pictures in blue or red were alway correct, for Group 2 (color-relevant) either blue or red was always correct (whether "equal-" or "unequal-number").

Group 3. In Group 3 the stimuli were identical on both days and insofar as the number of objects presented in the pair of pictures on each trial was concerned the stimuli were identical with those for Groups 1 and 2. But, in this group all objects were presented in black on a white background.

A different randomization of stimuli was used for every child in Group 1 and was matched across Groups 2 and 3. The randomization was highly restricted—each block of six trials including three "unequal-number" stimuli made up of one stimulus of the 1–2 or the 1–3 combinations of objects, one of the 2–1 or 2–3 combinations and one of the 3–1 or 3–2 combinations.

Procedure

To establish rapport, *E* spent several pre-experimental session hours with the children in their classroom. When taking each child to the experimental room to "play a game," *E* made use of a prepared list of questions and comments.

The child, in the experimental situation, was seated on a kindergarten-size chair before a kindergarden-size table on which stood the apparatus. The *E* when seated on a low chair on the other side of the apparatus was hidden from *S*'s direct view. The instructions—the verbalized aspects of which were minimized, the demonstrations maximized—consisted largely of showing the child some sample pairs of pictures of animals and birds with no more than one object in each of the sample pictures—and, demonstrating what happened when he pushed the "correct" or the "incorrect" button. For all *S*s in the pre-task demonstration, the response to the first stimulus was arbitrarily determined "correct" to the second stimulus, "incorrect." Following the "incorrect" response *S* was required to make a correction response. For the next two pre-task trials half the children were "correct" on Trial 3 and "incorrect" on Trial 4, with the reinforcement contingencies reversed for the remaining children.

After the demonstration of the procedure, the child was taken through 48 successive trials (each trial took approximately 5 seconds) or to criterion of 18 successive contact responses. On Day 2 after ensuring that *S* recalled the experimental procedures of button pressing and light reinforcement, he was taken through a further 48 trials or to criterion with Day 2 stimuli. Overt correction procedure was used throughout. In no

case did a child refuse to "play" or become inattentive during the experimental sessions.

RESULTS

Day 1. Thirty-five of the 48 Ss met criterion on Day 1. There were 11 learners in Group 1 and 12 in each of Groups 2 and 3. The mean number of errors for the learners in each of the Groups was 1.64, 2.00, and 1.62 for Groups 1, 2, and 3, respectively. Only those Ss who met criterion on Day 1 were subsequently taken through the Day 2 procedure.

Day 2. Group 1: Nine of the 11 learners reached criterion when number was relevant and color irrelevant. None of the nine Ss made any errors. Group 2: Of the 12 learners none reached criterion when color

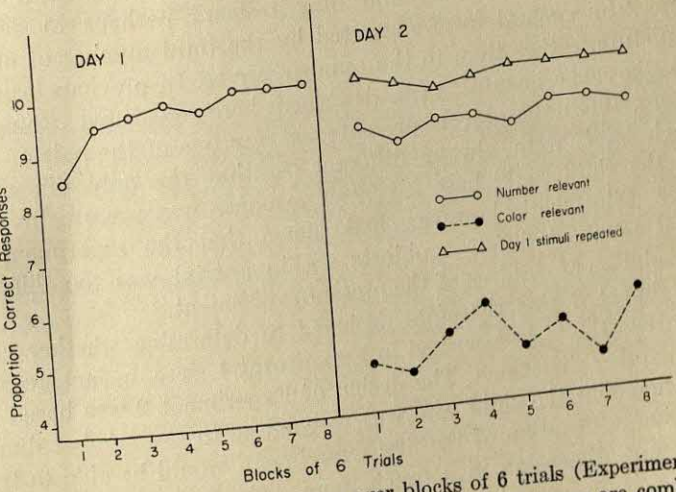


FIG. 1. Proportion of correct responses over blocks of 6 trials (Experiment 1) for the 35 Ss who met criterion on Day 1. The data for the 3 groups are combined on Day 1.

was relevant and number was irrelevant. The mean number of errors for the 12 learners on Day 2 was 23.83. Group 3: All 12 learners reached criterion on Day 2, on the same stimuli as on Day 1 with .32 mean errors. The proportions of correct responses over blocks of six trials for the learners in the three groups are presented in Fig. 1 with the data from the three groups combined for Day 1. On Day 2 there was no significant difference between the number of correct responses for Groups 1 and 3, but for Group 2—that is the groups with color-relevant and number-irrelevant—the probability of a correct response remains at the 0.5 level throughout the 48 trials.

DISCUSSION

The unequivocal findings in the present experimental situation, that with both number- and color-relevant, five-year-olds attended to the number and not the color dimension, was quite unexpected. The assumption that number is not a salient dimension for the young child is pervasive and although experimental evidence is slight it does provide some support (Lee, 1965).

It is important therefore, to ask whether in the present experiment, the child can have been responding correctly on the basis of any other dimension than number, and in fact, there seems to be some rationale for suspecting that this may have been the case. On each trial the stimulus presented to the child (two pictures) was made up of the same objects. Where the objects within a trial are the same, there seems some possibility that the child may respond perhaps to density or perhaps to a general form suggested by the total number of objects in each picture, rather than to the number *per se*. In previous experiments where the experimental task described here was used (Suppes and Ginsberg, 1963) the objects presented in any one of the pair of pictures on each trial, were always different, so that the only dimension on which the child could base a correct response was presumably that of number. Pilot work however, had shown that the experimental task when it involved different objects within a trial was too difficult for the population of children of the present experiment.

Experiment 2 was therefore designed to determine whether children who learned the task described in Experiment 1 were, in fact, responding to the number dimension. The design of Experiment 2 was based on the assumption that the child who was responding to the number dimension in the simple situation of the first experiment, would be able to transfer to the same dimension in a much more difficult situation. This line of reasoning is in accord with the "easy-to-hard" transfer phenomenon demonstrated by Lawrence (1952) and confirmed by others (Baker and Osgood, 1954).

EXPERIMENT TWO

Subjects

The *Ss* were 30 kindergarten children from two public schools. There were two groups, 18 *Ss* in Group 1, and 12 in Group 2, with an equal number of children from each school in each group.

Stimuli

The stimuli for Group 1 on Day 1 were those described for Group 3 in Experiment 1, that is they were made up of pairs of pictures, each

picture with from one to three objects. The objects were the same within a trial but differed across trials and all objects were black on a white background.

The stimuli for Group 1 on Day 2, and for Group 2 were the same. They were made up of pairs of pictures with each picture consisting of from one to three objects, but in this case all the objects in any one picture were different from each other. Eight different stamps were used to make the stimuli—a circle, triangle, rectangle, stick man, star, cross, and box. The height of the two new objects (stick man, and box) were 1½ inches and ½ inch, respectively. Sixty different pairs of pictures were prepared and over all stimuli each object appeared approximately equally often. A different sequence of stimuli was used for each child and the highly restricted randomization described in Experiment 1 was used in preparing each sequence.

Procedure

The same apparatus and the same detailed experimental procedures were used as described for Experiment 1. On Day 1 the 18 *S*s in Group 1 were taken through 48 trials or to criterion on the simple Day 1 stimuli. On Day 2 the learners from Day 1 and all *S*s in Group 2 were taken through 60 trials or to criterion on the more difficult stimuli. Sixty trials were used with the more difficult stimuli because the task had been found in previous pilot work to be extremely difficult for kindergarten children in schools in the same district—the majority of the pilot children failing to reach criterion in 60-trial sessions on each of three successive days.

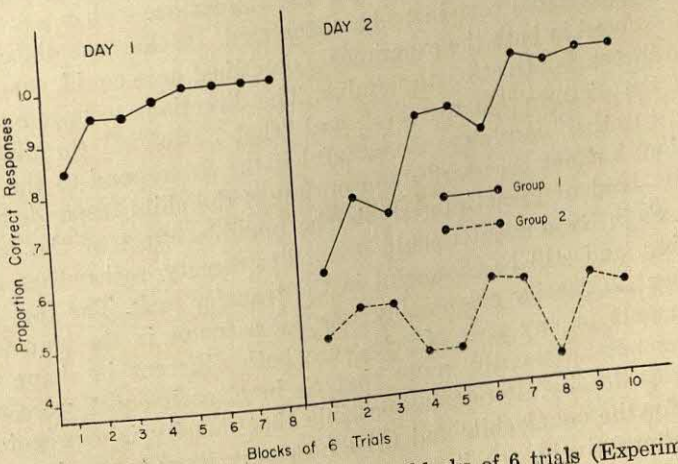


FIG. 2. Proportion of correct responses over blocks of 6 trials (Experiment 2).

RESULTS

Of the 18 *Ss* in Group 1, 15 met criterion on Day 1 and of these 15 all met criterion on Day 2 on the more difficult stimuli. Of the 12 *Ss* in Group 2, none met criterion in 60 trials on the more difficult stimuli. Figure 2 presents the proportion of correct responses over blocks of six for the two groups.

From Fig. 2 it can be seen that whereas the probability of a correct response in Group 2 remains around 0.5 throughout the 60 trials—in Group 1, by trial 24 the probability of a correct response is better than 0.9.

DISCUSSION

Again the experimental results are unequivocal—successful transfer to the task in which, as far as one can see, number is the only relevant dimension, suggests convincingly that *Ss* were, in fact, attending to the number dimension in the pretransfer trials. And if number is the pertinent dimension in Experiment 2 it seems reasonable to conclude that *Ss* in a similar situation in Experiment 1 responded to the same dimension.

The findings that *Ss* in Experiment 1 learned number rather than color when both dimensions were relevant are not entirely congruent with Lee's (1965) results. Lee's *Ss* made more errors in utilizing number than color, but Lee of course used only five trials per concept as compared to the long series of learning trials in the present experiment. More importantly, Lee's experimental task involved simultaneous discrimination the present task, successive discrimination. Ongoing experiments in this laboratory indicate that the results of the present experiment are proper to successive, but not to simultaneous discrimination.

Two other factors remain to be considered. In the first place, the instructions used in both the experiments described here could, conceivably, tend to direct *Ss* attention to number. The directions to "press this button" if the pictures "go together" and "that button" if the pictures do not "go together" is always meaningful if the *Ss* respond to the number dimension, but not necessarily appropriate if the child responds to color. Secondly—and of great interest—is the possible effect of the factor of density. *Ss* in Experiment 1 could respond to density, rather than to color or number, and still be successful in the transfer task. The fact that in Experiment 2, density presumably was not a factor in the transfer task (the elements in any pair of sets were both different in shape and in size) does not necessarily imply that *Ss* in Experiment 1 were not responding to density. It remains possible that density is a prepotent dimension for the young child and (a) a necessary stage to number acquisition or, alternatively, (b) it may facilitate number acquisition.

In any case whether density is or is not a prepotent dimension—an interesting possibility, but one which the present experiments do not explore—it remains clear from the evidence presented that in the appropriate situation, number can be an unexpectedly salient dimension for the five-year-old. Perhaps the fact that number is commonly observed to be a difficult dimension for the young child reflects not the young *Ss'* inability to deal with the principle involved, but rather the inappropriate nature of the task.

REFERENCES

BAKER, R. A., AND OSGOOD, S. W. Discrimination transfer along a pitch continuum. *Journal of Experimental Psychology*, 1954, **48**, 241–246.

BRIAN, C. R., AND GOODENOUGH, F. L. The relative potency of color and form perception at various ages. *Journal of Experimental Psychology*, 1929, **12**, 197–213.

KAGAN, J., AND LEMKIN, J. Form, color, size in children's conceptual behavior. *Child Development*, 1961, **32**, 25–28.

LAWRENCE, D. H. The transfer of a discrimination along a continuum. *Journal of Comparative and Physiological Psychology*, 1952, **45**, 511–516.

LEE, L. C. Concept utilization in preschool children. *Child Development*, 1965, **36**, 221–227.

PIAGET, J. How children form mathematical concepts. *Scientific American*, 1953, **189**, 74–79.

SRA Series, Chicago, Ill., Science Research Inc., 1961.

SUCHMAN, R. G., AND TRABASSO, T. Color and form preference in young children. *Journal Experimental Child Psychology*, 1966(a), **3**, 177–187.

SUPPES, P. *Sets and Numbers*, L. W. Singer Company, 1963.

SUPPES, P., AND GINSBERG, R. Experimental studies of mathematical concept formation in young children. *Science Education*, 1962, **46**, 230–240.

SUPPES, P., AND GINSBERG, R. A fundamental property of all-or-none models, binomial distribution of responses prior to conditioning, with application to concept formation in children. *Psychological Review*, 1963, **70**, 139–161.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 274–281 (1969)

Cross-modal Transfer of a Discrimination by Retarded Children[1]

JEROME SMITH AND JEFFREY TUNICK
University of Connecticut

A miniature experiment design was employed to test transfer of a discrimination from visual- to tactual-kinesthetic (active touch), and tactual-kinesthetic to visual sense modes. Cross-modal transfer was exhibited by retarded *Ss* when cues were identical for both modalities. When only information about the dimension relevant to task solution was available, no cross-modal transfer was demonstrated.

The ability to transfer information gained through one sense mode to problem solution requiring the use of another sense mode has been termed cross-modal transfer. The magnitude of such transfer has been found to relate to phylogenetic level, type of task, and in the human to developmental level as well as type of task.

The role of phylogenetic level has been demonstrated by the failure to find cross-modal transfer in monkeys by Ettlinger and his associates (Ettlinger, 1960; Burton and Ettlinger, 1961) and the presence of cross-modal transfer in humans demonstrated with children by Birch and Lefford (1963) and Blank and Bridger (1964) and with college students by Tyrrell (1967). That type-of-task as well as phylogenetic level must be considered is indicated by the presence of cross-modal transfer in monkeys when the discrimination problem in both sense modes involves identical cues (Wilson and Schaeffer, 1963).

Developmental effects have been revealed by a correlation between age and performance on a cross-modal recognition task in the Birch and Lefford study. A similar correlation for a recognition task was found in the Blank and Bridger study. Type of task also plays a role in determining the extent of cross-modal transfer with human subjects. The children in Blank and Bridger's study who had successfully transferred recognition across modes experienced great difficulty with a concept-transfer task. The authors conclude that recognition task may not de-

[1] This study was supported by grants from the University of Connecticut Research Foundation and from the U. S. Office of Education, Department of Health, Education and Welfare.

pend upon verbal mediation while concept-transfer tasks do. The role of verbal mediation is stressed by Ettlinger (1960), and Hermelin and O'Connor (1961) as well. Indeed Hermelin and O'Connor point to the failure of their retarded subjects to transfer recognition of *unfamiliar* forms cross-modally as evidence for the need for verbal labels.

While verbal mediation has been successfully used as an explanatory concept in the discrimination learning of children (Kendler and Kendler, 1959), other mediational concepts have also been proposed. Zeaman and House (1963), for example, have described an Attention Theory which specifies a two-stage process of retardate learning: A first step in which an observing response to a stimulus *dimension* is made, and a second step in which an instrumental response to a specific *cue* within the dimension is made. The term *cue* refers to a specific characteristic of one of the discriminanda. In a two choice situation requiring the choice of triangle as opposed to a square, *square* and *triangle* are cues. Both square and triangle, however, have something in common, they are members of a broader class which might be called geometric *form*. Zeaman and House regard form as a dimension. "Dimensions are broad classes of cues having a common discriminative property" (Zeaman and House, 1963). While the isolation of psychological dimensions is a psychophysical task which is far from complete, the gross distinctions made by Zeaman and House have proven valuable in retardate visual-discrimination-learning (Zeaman and House, 1963; House and Zeaman, 1963) and in the tactual-kinesthetic discrimination of retardates as well (Brown and Smith, 1967). Indeed, mediation through an observing response has provided a useful alternative to verbal mediation (Shepp and Turissi, 1967). The present study, therefore, is designed within the framework of observing response mediation.

The question posed is whether the retarded child can transfer across sense modes information about specific cues, stimulus dimension or both? On a theoretical level it is of interest to discover whether retardates, who can transfer dimensional and cue information within a sense mode (Zeaman and House, 1963) can do so between sense modes. If retardates fail to transfer any information across modes, the generality of Attention Theory may be questioned. On an empirical level, further specification of the amounts and types of cross-modal transfer in a retarded population should prove profitable on two counts: (1) Providing more information about cross-modal transfer in general, and (2) possibly pointing to an area of retardate deficit worthy of more extended investigation.

The present study examines the performance of retarded children on the transfer of a two-choice discrimination from the visual to the tactual-kinesthetic (active touch) modes and conversely from the tactual-

kinesthetic to the visual. The independent variables of interest are the direction of transfer, whether the information can be transferred in terms of dimensions or cues, or both, and the kind of dimension relevant to task solution. Since the conditions resulting from the number of variables investigated and the appropriate control conditions are large (12), a miniature experiment design (House and Zeaman, 1963) is employed. The miniature experiment uses each *S* as his own control, a small number of trials per problem and many problem repetitions, providing an efficient technique for the study of several variables within the same experimental situation.

METHOD

Subjects. *Ss* were twelve institutionalized retardates. Six of these were unable to meet pretraining criteria leaving an *N* of six males for the experiment proper. Mean MA, CA, and IQ were 85 months, 147 months, and 58, respectively.

Apparatus. *S*'s were run on a modified Wisconsin General Test Apparatus. The apparatus included an enclosed chamber which was pushed toward the child for a tactual-kinesthetic (TK) problem. Two circular openings covered by foam rubber flaps permitted insertion of both hands but no visual inspection. Two objects were manually explored simultaneously and the chosen one lifted by *S* to gain access to a reinforcement well beneath the object. For visual problems, the bottom half of the chamber containing the reinforcement wells was pushed toward *S* permitting visual inspection of the stimulus objects. On visual trials both stimuli were enclosed in clear plexiglas wedges to prevent differential TK cues accompanying visual cues.

The discriminanda were plastic equal volume geometric forms. All combinations of six forms (cone, sphere, pyramid, cube, rectangular solid, and cylinder) and four textures (smooth, foam rubber, buck-shot, and rough sandpaper) were used, providing a pool of 24 stimulus objects.

Procedure

In all phases of the experiment correct responses were reinforced with an M & M candy and a verbal "good" from *E*. Incorrect responses were followed by a "no" from *E*. Non-correction procedure was used throughout.

Pretraining. *Ss* were required to solve eight pretraining problems before entering the experiment proper. Each problem had to be solved in a *single* daily session of 25 trials with a criterion set at 20 of 25 correct. If more than one day was required to reach criterion on any given problem, *S* was presented with a similar problem. If criterion was not reached for

any given problem type within 125 trials *S* was excluded from the experiment. If, after reaching criterion twice, single day solution was not achieved, *S* was dropped.

The pretraining problems included: one visual junk and one TK junk problem, one visual problem using the experimental objects with *both* texture and form relevant, a TK problem of the same type, two visual problems one with form relevant and one with texture relevant, and two TK problems of the same type. Irrelevant dimensions were variable in all problems.

Testing. A miniature experiment technique utilizing each subject as his own control and requiring the administration of several problems each day was employed. Each *S* was given 12, four-trial problems daily for a period of 30 days. There were 12 problem types and each of the 6 *S*s received 30 repetitions of each problem type, providing a total of 180 measures for each trial of each problem. All 12 problem types were presented in a daily session in a random order. After the fourth trial of each problem *S* was told: "We have finished this part of the game and we are going to start another part." Six random orders were presented five times each over the 30-day testing period.

Before beginning a daily experimental session each *S* had to reach a criterion of 7 correct consecutive responses in 15 trials of warm-up problems. Two warm-up problems, one visual and one TK, with both texture and form relevant were presented. If *S* failed to reach criterion on warm-up, he was excused from the session and brought back the next day.

The 12 problem types consisted of 8 experimental and 4 control problems. The variables manipulated were: Type of transfer (dimensional, cue, or control), direction of transfer (visual to TK vs TK to visual), and dimension relevant to solution (texture or form). Table 1 presents the problem types and Table 2 provides examples of the four-trial problem.

Table 2 shows that experimental problems required two trials in one sense mode and two in the other, while the control problem presented all four trials in one sense mode only. The cue problems utilize the same cues in both sense modes while the dimensional problems change the cues for the third and fourth trials. The only information available for transfer across modes in the dimensional problem is which dimension is relevant to solution. Relevant cues for any single problem were randomly selected with the restriction that the cues of the first two trials of any problem differed from the last two of the preceding problem. The irrelevant cues were selected in similar fashion. The position of the correct cue and the positions of the irrelevant cues were randomly assigned with the restriction that no sequence of four consecutive rights or lefts was permitted.

TABLE 1
EXPERIMENTAL DESIGN AND NUMBER CORRECT RESPONSES ON TRIALS THREE AND FOUR

Problem type	Type of transfer	Direction of transfer	Relevant dimensions	Trial 3	Trial 4
I	Dimension	Visual to T-K	Form	(93)	(91)
II			Texture	(91)	(111)
III		T-K to Visual	Form	(87)	(77)
IV			Texture	(97)	(102)
V	Cue	Visual to T-K	Form	(132)	(129)
VI			Texture	(140)	(147)
VII		T-K to Visual	Form	(124)	(123)
VIII			Texture	(149)	(137)
IX	Visual control	None_____	Form	(134)	(140)
X		None_____	Texture	(147)	(153)
XI	T-K Control	None_____	Form	(131)	(121)
XII		None_____	Texture	(148)	(143)

RESULTS

Since the irrelevant dimensions were variable within trials, learning is reflected in the performance on Trials three and four only. The number correct for each problem type on Trials three and four is presented parenthetically in Table 1. Since 6 *S*'s were given 30 presentations of each problem type, the maximum number for each trial is 180. A factorial analysis of variance ($3 \times 2 \times 2 \times 2 \times 6$) was performed on the response measure of number correct. The factors included problem type (dimen-

TABLE 2
ILLUSTRATIVE PROBLEMS

Trial	Dimension (I)		Cue (V)		Control (XI)	
	Sense mode	Cues	Sense mode	Cues	Sense mode	Cues
	V	+ -	V	+ -	TK	+ -
	V		V		TK	
	TK		TK		TK	
	TK		TK		TK	

sion, cue, or control), sense mode (V or TK), relevant dimension (form or texture), trials (3 or 4) and subjects. A significant effect of problem type $F = 36.25$, df 2,10, $p < .001$ was found as well as a preference for texture as opposed to form relevant ($F = 12.31$, df 1,5, $p < .025$). No effects of sense mode which represented the direction of transfer (V to TK, or TK to V) was found, nor was the performance on Trials three and four significantly different. None of the interactions approached significance.

Since no effects of transfer direction were found and since the dimension-relevant differences were not accompanied by a significant interaction of that factor with any of the others, these variables along with the subject variable were collapsed to provide the function of problem type by trials presented in Fig. 1. Number correct has been converted to percent correct in this figure.

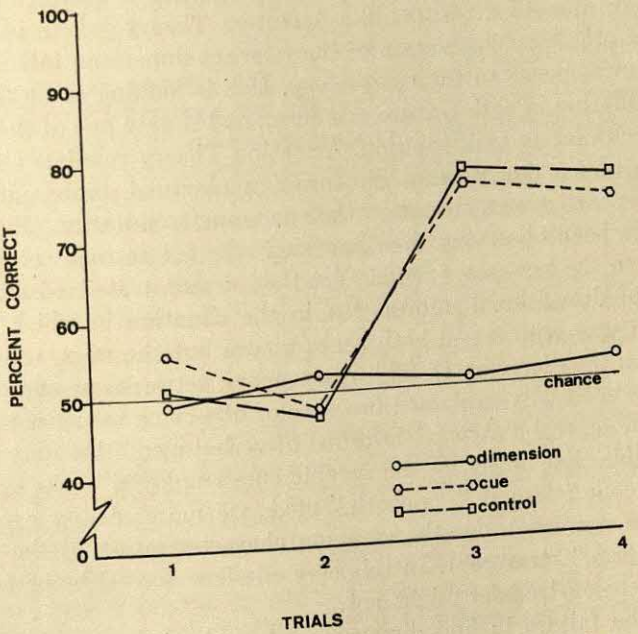


FIG. 1. Percentage correct for dimensional transfer, cue transfer, and control problems.

Figure 1 reveals that performance on the dimensional problem remains close to chance throughout all four trials while both the cue-transfer and control problems rise to almost 80% on trial three and remain there for trial four.

Orthogonal comparisons among the levels of the problem type factor reveal a significant difference between the cue-transfer and control prob-

lems on the one hand and the dimensional-transfer on the other ($F = 518$, df 1,10, $p < .001$) but no difference between the cue and control problems ($F = 4.1$, df 1,10, $p > .05$).

DISCUSSION

The evidence for cross-modal transfer of a discrimination by retardates appears clear-cut. When the same cues are involved (cue-transfer) the transfer of problem solution across modalities is as good as performance in control problems which require the use of the same sense mode throughout. When only information about the dimension relevant to task solution is available (dimension-transfer), no evidence for transfer is obtained. In light of the high level of cross-modal transfer for like-cue situations this lack of dimensional transfer is surprising. The question of whether the retarded child can transfer dimensional as well as cue information is of importance to a theory, like Attention Theory, which postulates a two stage process: Observation of the relevant dimension followed by instrumental response to the correct cue. If *S* is looking at or feeling two objects differing in both texture and form, and if only one of these dimensions is relevant to task solution, Attention Theory predicts chance performance unless the relevant dimension is observed. Since our *S*'s were performing well above chance in the cue transfer situation (Fig. 1) they must have been observing the dimension relevant to task solution when switched to the new sense mode. Yet they demonstrated no cross-modal transfer of dimensional information in the situation in which the same dimension was relevant in both sense modes but the cues were changed (dimension transfer, Fig. 1). This discrepancy between cue and dimension transfer is not easily explained in terms of observing response mediation, although a general statement in terms of verbal mediation may be made. It is possible that labels for the specific cues employed in the experiment ("ball," "cone," "rough," "smooth," etc.) are more readily available to the *S* than are labels for the common characterization of the discriminanda ("form," "texture"). In this case cue-transfer might be facilitated while dimension-transfer might not.

While the failure to find dimensional transfer is not easily explained by Attention Theory as we have described it, it can be handled within the Zeaman and House theoretical structure. The theory is composed of several mathematical models of two-choice discrimination, each differing to some extent in the basic assumptions about how the retarded child deals with a two-choice problem. The primary model assumes that the child observes only *one* dimension at each trial. It is this "one-look" model which we have applied to our data, and which is faulted by the failure to find dimensional transfer in the presence of a high level of cue

transfer. Suppose, however, that the child can observe both texture and form on each trial and does so regularly. Providing dimensional information without cue information would prove of little value. Yet, this was the situation in the dimension-transfer condition. A "multiple-look" model has been proposed by Zeaman and House (Zeaman and House, 1963; House and Zeaman, 1963) which does permit observation of more than one dimension on each trial. Such a model may be more appropriate for our data, especially since *S*'s were given extensive pretraining in both texture and form observation to facilitate performance in the four-trial experimental problems.

Retarded children can transfer discriminations across sense modes. Whether they can transfer information in terms of dimensions, however, is still an open question. A conventional trials-by-subjects design may provide the answer, and would appear worth pursuing.

REFERENCES

Birch, H., and Lefford, A. Intersensory development in children. *Monographs of the Society in Child Development*. 1963, **28**, 5.

Blank, M., and Bridger, W. Cross-modal transfer in nursery school children. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 277–282.

Brown, Z., and Smith, J. Tactual-kinesthetic discrimination shifts in retarded children. *Journal of Experimental Child Psychology*, 1967, **5**, 455–462.

Burton, D., and Ettlinger, G. Cross-modal transfer of training in monkeys. *Nature*, 1961, **191**, 308.

Ettlinger, G. Cross-modal transfer of training in monkeys. *Behaviour*, 1960, **16**, 56–65.

Hermelin, B., and O'Connor, R. Recognition of shapes by normal and subnormal children. *British Journal of Psychology*, 1961, **52**, 281–284.

House, B. J., and Zeaman, D. Miniature experiments in the discrimination learning of retardates. *In* L. P. Lipsitt and C. C. Spiker (Eds.) *Advances in Child Development and Behavior* (I). New York: Academic Press, 1963.

Kendler, T. S., and Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **85**, 56–60.

Shepp, B. E., and Turrissi, F. D. Learning and transfer of mediating responses in discriminative learning. *In* N. R. Ellis (Ed.), *International review of research in mental retardation*. Vol. 2. New York: Academic Press, 1966. Pp. 85–121.

Tyrrell, D. J. Cross-modal transfer of oddity learning. *Psychonmic Science*, 1967, **9**, 353–354.

Wilson, W. A., and Schaffer, O. C. Intermodality transfer of specific discriminations in the monkey. *Nature*, 1963, **197**.

Zeaman, D., and House, B. J. The role of attention in retardate discrimination learning. *In* W. R. Ellis (Ed.) *Handbook of Mental Deficiency*. New York: McGraw-Hill, 1963.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 282–298 (1969)

Sucking in the Newborn during a Feed

JUDITH DUBIGNON AND DUGAL CAMPBELL[1]

Queen's University, Kingston, Ontario

Two components of sucking (suction and expression) were measured under three conditions: when no fluid was available, when small amounts of dextrose were delivered, and throughout a milk feed. Significant differences in sucking rate, time spent sucking, and pressure were found between the nonnutritive and dextrose conditions; these differences were more marked during the feed.

Sucking throughout the feed remained uniform until the end of the last quarter when a reduction in frequency of sucking, time spent sucking, and sucking pressure took place. There were no changes in sucking rate. Heart rate rose during nutritive sucking and moment-to-moment variation in heart rate was reduced.

Experiments intended to identify the stimuli that control sucking have shown that intraoral stimuli are significant determinants of sucking in the newborn. For example, characteristics of the nipple, such as its size, shape, and compressibility, have been found to control the frequency of nonnutritive sucking (Lipsitt and Kaye, 1965; Dubignon and Campbell, 1968a). These results resemble those found in ethological studies that have shown that feeding behavior in the young of certain species is controlled by relatively specific sign-stimuli (Tinbergen, 1951).

Little is known about stimulus control of sucking during feeding; however, among the stimuli that control sucking in a functional context, it is likely that those provided by nutritive fluids are relatively important, as well as other factors, such as satiation and fatigue. Until recently "there have been no reports comparing sucking for liquids and nonnutritive sucking within the same recording system" (Kaye, 1967). In one such study the data suggested that nutritive sucking during a feed is different from nonnutritive sucking. It was found that the delivery of small amounts of dextrose brought about immediate changes in pattern

[1]The work reported in this paper was supported by the Ontario Mental Health Foundation (Grant No. 16) and the Association for the Aid of Crippled Children. We wish to record our thanks to the nursery staff of the Kingston General Hospital and to our research assistants, Miss M. Henderson, Mrs. E. Stephens, and Miss Elizabeth Wood.

of sucking: the infants sucked more slowly and gave longer bursts of sucking. Naturalistic observations of breast and bottle feeding suggested that these changes in pattern of sucking became even more marked during feeding. In the same study significant changes in heart rate were found to accompany sucking (Dubignon and Campbell, 1968b).

The present experiment was designed to examine further the hypothesis that the stimuli provided by nutritive fluids are significant determinants of sucking. The earlier observations led to the prediction that the differences found between nonnutritive sucking and sucking for small amounts of dextrose would become more marked during a feed. Secondary aims were to provide data on the regulation of food intake by the newborn, the effects of fatigue and satiation on nutritive sucking, and changes in heart rate during feeding.

METHOD

Sucking was measured under three conditions: first, nonnutritive sucking when no fluid was given to the baby; second, when small amounts of dextrose fluid were delivered; and third, throughout a bottle feed. The order of nonnutritive and dextrose conditions was not varied as it has been shown that the order of trials has no significant effect on the sucking elicited in each condition (Dubignon and Campbell, 1968b).

Subjects. Twenty-four full-term babies (12 male, 12 female) were seen during the third or fourth days of life (mean age: 70 hours; *SD* 12.1). All the babies were bottle-fed infants of multiparous mothers, and had no known or suspected pathological condition.

Apparatus. The apparatus gave measures of sucking behavior and heart rate from infants held in the ordinary feeding position. The device used to measure sucking (Fig. 1) was designed to record two components of sucking behavior: *expression,* or the positive pressure created when the baby approximates his gums and elevates his tongue, and *suction,* or negative pressure produced when the baby lowers his tongue and bottom jaw. Pressure changes produced by suction and expression were monitored by two pressure transducers and written out on two channels of a polygraph using DC preamplifiers. Before each experimental session, the apparatus was calibrated by applying standard negative and positive pressures to the nipple.

During the feed a rotary pump moved the milk along the delivery tube. The action of the pump was controlled by a trigger circuit fired by the polygraph; for half of the group (12 infants) the pump was triggered by expression responses and for the other half by suction responses. The aim of the experiment was to record normal sucking behavior during a feed that closely resembled a natural bottle feed. Accordingly, the

minimum pressures that triggered the pump (—20 mm Hg for suction and 4 mm Hg for expression) were set very low to avoid influencing normal suction or expression pressures by reinforcement contingencies. The pump delivered 0.16 cc of milk each time it moved so that the infant had to suck 190 times in order to get 1 ounce of milk.

Sucking was also recorded by counting the movements of the mouth and jaw made by the baby. An observer signalled every tenth count with an event marker on the polygraph.

An EKG was taken from electrodes on the baby's chest and back. These led to two channels of the polygraph that recorded both the electrocardiogram and a tachograph tracing of heart rate.

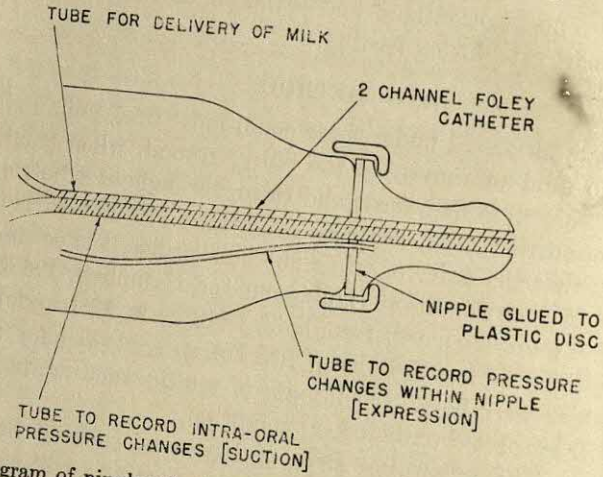


FIG. 1. Diagram of nipple adapted to deliver fluid and to measure the suction and expression components of sucking.

Procedure. At a normal feeding hour the subject was brought to the research room. The nursery regimen called for a feed every 4 hours. Diapers were changed in the nursery to provide a standard rousing procedure. The baby was weighed, heart rate electrodes were applied, and the baby was bundled in a carrying blanket and given to the first experimenter (*E*1). *E*1 held the baby in the normal feeding position, i.e., in the crook of the left arm with the head resting against the experimenter's chest. *E*1 was responsible for giving the baby the nipple, signalling: (i) the different phases of the experiment and, (ii) every tenth suck during the sucking trials. A second experimenter (*E*2) operated the polygraph, timed the trials, delivered dextrose through the nipple during the second trial, and prepared the milk delivery system for the feed.

At the start of the experiment, the baby was held by *E*1 in the feeding

position for 2 minutes. During this time only a heart-rate record was made. Immediately afterwards the baby was given a nipple, and once this had been accepted, 2 minutes of nonnutritive sucking were recorded. After this first trial, a syringe filled with a 5% dextrose solution (temperature approximately 35°C) was attached to the delivery tube of the catheter so that during the second trial *E*2 delivered .5 cc of dextrose for every tenth suck counted by *E*1. The dextrose trial followed the nonnutritive trial by approximately 30 seconds and also lasted 2 minutes.

During the third part of the experiment, expression, suction, and heart rate, were recorded while the baby was given its usual milk feed by means of the pump. *E*1 held the baby as before and signalled: (i) presentation and acceptance of the nipple, (ii) her count of sucking and, (iii) withdrawal of the nipple when burping was necessary. During burping periods the polygraph was switched off and the length of these periods was noted on the record before the feed was resumed. When the baby rejected the nipple, the bottle was removed from the pump-delivery mechanism; the amount of milk consumed and time taken over the feed was noted. The infant was offered what remained of the feed directly from the bottle, heart rate was recorded, and *E*1 continued to count sucks if the baby sucked at all at this point. When the nipple was finally rejected, any additional amount taken was noted and heart rate was recorded during a final 2-minute rest period. The baby was weighed and returned to the nursery.

Scoring and analysis. Scoring criteria were calculated for each chart so that every record could be scored in terms of a standard preamplifier sensitivity setting on the polygraph and a "standard" nipple. The minimum standards for scoring pressure changes corresponded to a pressure of —20 mm Hg for suction and 4 mm Hg for expression (cf. Dubignon, 1967).

The first 2 minutes of the feed were treated as a 2-minute trial comparable to the non-nutritive and dextrose trials. The last 90 seconds of each 2-minute trial were scored to avoid the initial burst of sucking at the beginning of each trial. The following scores were made for both expression and suction: (i) a count of the number of responses above criterion amplitude; (ii) time spent responding during the 90 seconds (the criterion for time out from responding was any break in the record longer than 1 second); (iii) rate (calculated from the first two measures by dividing the time spent responding into the count); (iv) the amplitude of every fifth response; (v) expression/suction ratio (calculated by dividing the suction count into the expression count); and, (vi) the observer's count for each 2-minute trial was recorded.

Heart rate was scored for the last 45 seconds of each 2-minute trial and heart-rate variability was measured by recording the difference between the highest and lowest point on the tachograph tracing for the same period. This measure is referred to as heart-rate range.

Sucking during the whole feed was first studied by dividing each pump-feed record into four sections of equal duration (excluding burping periods). The end of the pump feed was taken as the last scorable response. The last 45 seconds of each quarter section of the feed were scored for suction and expression responses and the same six measures were taken as in the three earlier trials. Heart rate and heart-rate range were scored for the same four periods with the addition of a further score for the last 45 seconds of the final rest period.

The influence of fatigue on sucking during the feed was examined by scoring additional 45-second sections of the records before and after the first and last burping periods; this was done for half of the group ($N = 12$).

The results were treated by analysis of variance and evaluated as mixed models. An arcsin transformation was applied to the time scores before analysis.

Reliability of scoring. The reliability of polygraph scores was examined by having ten records scored independently by two scorers ($r = 1.0$ for counts, $r = .99$ for time spent sucking, and $r = .99$ for amplitude scores). The method of scoring amplitude was checked for reliability by taking ten records and scoring every response. The mean amplitudes obtained in this way were correlated with the means obtained from scoring every fifth response ($r = .97$).

RESULTS

Comparisons of pump feeding with natural bottle feeding. The total duration of the pump feed, the number of sucks counted by the observer, and the number of rest periods for burping were recorded for each of the 24 infants. These data were compared with equivalent figures for 20 infants given bottle feeds by their mothers in a previous experiment (Dubignon and Campbell, 1968b). No significant differences were found between pump and natural bottle feeding although a natural feed took a shorter time, required fewer burping periods, and led to a lesser number of sucks on the average.

Sucking scores. The over-all mean and range of the different measures of sucking behavior are summarized in Table 1. These scores fall within the range of scores reported by other investigators (Balint, 1948; Colley and Creamer, 1958; Halverson, 1938, 1944; Kron *et al.*, 1967; Sameroff, 1965).

TABLE 1
MEAN AND RANGE OF SCORES FOR MEASURES OF SUCKING BEHAVIOR

	Mean	Range
Suction count/minute	49.77	5–87
Expression count/minute	53.58	14–87
Suction time/minute (seconds)	37.96	4–58
Expression time/minute (seconds)	40.16	13–59
Suction rate per second of sucking time	1.326	.951–1.976
Expression rate per second of sucking time	1.351	.898–1.867
Suction pressure (mm Hg)	−82.0	−(20–185)
Expression pressure (mm Hg)	23.5	5–48

Treatment effects. The mean scores for expression and suction under each treatment for all six measures are summarized in Table 2 together with F values. Duncan's Multiple Range tests were used to test the differences between the individual means.

Sucking counts. The introduction of dextrose on the second trial led to a sharp drop in expression count followed by a return to the nonnutritive level during the feeding trial ($p < .05$). The same effect was found in the observer's count ($p < .05$). The suction count varied in the same direction across trials but the effect was not significant.

TABLE 2
TREATMENT MEANS FOR EXPRESSION AND SUCTION ON EACH MEASURE

Measure	Non-nutritive trial	Dextrose trial	Feeding trial	F
Sucking counts (in 90 seconds)	79.4	70.3	74.2	1.55
Suction count	85.4	72.7	83.0	4.50*
Expression count				
Observer's count (for each 2-minute trial)	.941	.814	.901	3.60*
Sucks per second				
Time spent sucking (in 90 seconds)	50.7	55.8	64.3	7.28**
Suction time	54.9	56.3	69.5	14.06***
Expression time				
Sucking rate (per second of actual sucking time)	1.55	1.26	1.16	78.99***
Suction rate	1.54	1.30	1.21	56.94***
Expression rate				
Amplitude of responding (mm)	9.19	10.78	10.92	10.30***
Suction amplitude	9.91	11.19	8.93	2.50
Expression amplitude	1.17	1.05	1.16	—
Expression/suction ratio (over 90 seconds)				

* $p < .05$; ** $p < .01$; *** $p < .001$.

Time spent sucking. During the feeding trials the time spent in suction and expression was significantly higher than in either the non-nutritive or dextrose trials ($p < .01$).

Sucking rate. There was a marked decline in suction and expression rates during the dextrose trial ($p < .001$) and a further significant decline during feeding ($p < .01$).

Amplitude of sucking. Suction amplitude was significantly lower ($p < .001$) during non-nutritive sucking than in the dextrose and feeding trials in which the suction amplitudes did not differ. Amplitude of expression did not change significantly across trials.

Expression/suction ratio. There were no significant differences in the ratio of the two components of sucking during the three trials. The ratio was always greater than 1, showing that the expression component appeared more often than the suction component.

Trigger group effects. There were no significant differences on any measure between the scores obtained by the two groups for whom the pump was triggered by suction and expression respectively. Nor were any of the trigger group $\times$ treatments interactions significant.

In summary, the results show that non-nutritive sucking differed from sucking when nutritive fluid was delivered. Fluid delivery was accompained by a change in the pattern of sucking, i.e., by an increase in time spent sucking, slower sucking rates, and an increase in amplitude of suction. To a lesser extent sucking during a feed differed from nutritive sucking for small amounts of dextrose. This was shown by the marked increase in time spent sucking during the feeding trial that resulted in a higher sucking count in spite of the slower sucking rate. While a change in pattern of sucking occurred when nutritive fluid was delivered, it appears that the mechanism of sucking did not change, in that the relationship between the expression and suction components of sucking remained the same.

Sucking behavior during the feed. The mean scores for expression and suction during each quarter of the feed for all six measures are summarized in Table 3 together with F values.

Sucking counts. The counts of suction and expression responses were significantly lower ($p < .001$) during the last quarter of the feed than in the first three quarters in which the counts did not differ. The observer's count made throughout each quarter of the feed revealed the same effect ($p < .001$) and also showed that the number of sucks made during the third quarter was lower than during the first quarter ($p < .05$).

Time spent sucking. During the last quarter of the feed there was a significant drop in suction time ($p < .001$) and expression time ($p <$

TABLE 3

MEAN SCORES FOR EACH MEASURE OF EXPRESSION AND SUCTION DURING THE LAST 45 SECONDS OF EACH QUARTER OF THE FEED

Measure	First quarter	Second quarter	Third quarter	Fourth quarter	F
Sucking counts (in 45 seconds)					
Suction count	36.46	34.29	31.87	18.92	12.39***
Expression count	42.25	39.41	39.46	30.42	11.24***
Observer's count (for duration of each quarter)					
Sucks per second	.859	.817	.759	.568	24.79***
Time spent sucking (in 45 seconds)					
Suction time	30.04	28.08	26.44	15.90	10.86***
Expression time	34.63	32.46	32.39	25.92	6.01**
Sucking rate (per second of actual sucking time)					
Suction rate	1.22	1.22	1.15	1.23	1.01
Expression rate	1.25	1.22	1.23	1.17	1.23
Amplitude of responding (mm)					
Suction amplitude	9.57	9.55	8.61	6.18	11.93***
Expression amplitude	10.08	9.42	9.34	6.92	6.60***
Expression/suction ratio (over 45 seconds)	1.292	1.226	1.429	1.799	3.61*

* $p < .05$; ** $p < .005$; *** $p < .001$.

.005) as compared with the first three quarters in which time spent responding did not differ.

Sucking rates. There were no significant differences in suction or expression rates throughout the feed.

Amplitude of sucking. During the last quarter of the feed there was a sharp drop in amplitude of suction ($p < .001$) and expression ($p < .01$) as compared with the first three quarters in which amplitude of suction or expression did not differ significantly.

Expression/suction ratio. There was an increase in the ratio of expression to suction responses in the last quarter of the feed ($p < .05$).

Trigger group effects. Suction amplitude was slightly higher over-all in the group of subjects for whom the pump was triggered by the expression component of sucking. The mean amplitudes were: 7.21 mm for the suction trigger group and 9.75 mm for the expression trigger group ($F = 4.36$, df 1,22, $p < .05$). There were no significant differences between the two groups on any other measure.

In summary, these results show that the infants did not control their food intake by altering their sucking rate. Satiation was marked by less time spent sucking (and a consequent reduction in sucking count) and reduced amplitude of suction. In addition there appeared to be a change in the sucking mechanism with satiation as shown by the increase in expression/suction ratio. This was due to a marked drop in the frequency of the suction component during the last quarter of the feed.

Sucking behavior before and after burping. The mean scores for the main effects are shown in Table 4 together with F values. There were no significant interactions.

Before and after burping. The sucking counts for suction and expressions were higher after burping than before burping ($p < .005$), and there was a corresponding increase in suction time ($p < .025$) and expression time ($p < .005$). There were no significant differences before and after burping in suction or expression rates.

First and last burp. At the time of the first burp sucking counts for both expression ($p < .005$) and suction ($p < .025$), and time spent in suction ($p < .01$) were higher than at the last burp. Time spent in expression varied in the same direction but did not reach significance. There were no significant differences between sucking rates at the first and last burping periods.

In summary, these results show that after a burping period the infants increased the time spent sucking and showed a corresponding increase in sucking counts. The recovery after a rest was as great in the early part of the feed as toward the end of the feed. Toward the end of the feed (last burp) sucking counts and time spent sucking dropped off as

TABLE 4
SUCKING SCORES IN RELATION TO REST PERIODS

	Mean sucking scores before and after burping periods			
	Before	After	F	p
Suction count (45 seconds)	30.04	41.54	13.99	$<.005$
Expression count (45 seconds)	38.17	46.54	15.78	$<.005$
Suction time (45 seconds)	24.79	34.50	8.97	$<.025$
Expression time (45 seconds)	30.88	38.31	14.75	$<.005$
Suction rate (per second)	1.166	1.225	—	NS
Expression rate (per second)	1.257	1.223	—	NS
	Mean sucking scores at first and last burping periods			
	First	Last	F	p
Suction count (45 seconds)	42.21	29.38	9.53	$<.025$
Expression count (45 seconds)	45.92	38.79	18.11	$<.005$
Suction time (45 seconds)	35.23	24.06	11.29	$<.01$
Expression time (45 seconds)	36.42	32.77	3.11	NS
Suction rate (per second)	1.197	1.194	—	NS
Expression rate (per second)	1.277	1.204	1.50	NS

compared with the early part of the feed (first burp) but sucking rates did not change.

Heart-rate changes. The mean heart rate and heart-rate ranges scored over 45-second periods under the various conditions are plotted in Figure 2. Time spent sucking (expression) is shown at the top of the figure. For the quarters of the feed, time spent sucking in 45 seconds has been doubled to give scores comparable with those for the 90-second trials. Analyses of variance were carried out on the heart-rate measures for the eight conditions (excluding the scores for the feeding trial, which were identical to those for the first quarter of the feed). Highly significant differences were found between conditions ($p < .001$) for both heart rate and heart-rate range. Multiple range tests revealed that: (i) during feeding (all four quarters) heart rate was faster than during non-nutritive sucking or initial and final rest periods; (ii) during feeding (quarters two and three), heart rate was faster than during the dextrose trial; (iii) during the first three quarters of the feed heart-rate range was lower than during either rest period; and (iv) during the first half of the feed, heart-rate range was lower than during the last quarter of the feed or non-nutritive and dextrose trials.

In summary, these results show that during nutritive sucking heart rate increased sharply and moment-to-moment variation in heart rate

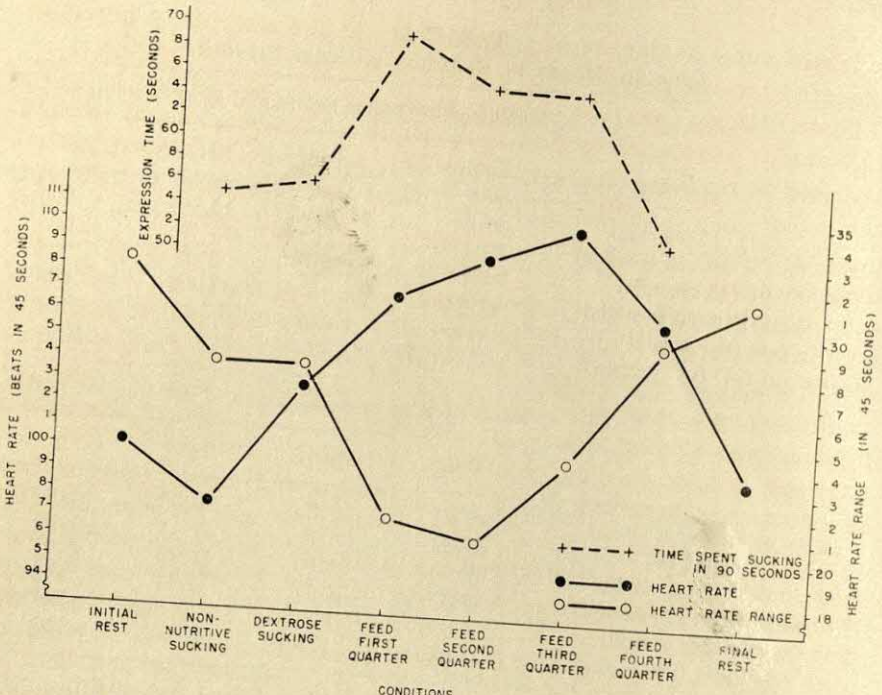


FIG. 2. Measures of heart rate as related to sucking and rest.

(heart-rate range) showed a corresponding decrease below the resting levels. The plot of time spent sucking, which has been superimposed on Fig. 2, makes it clear that the reduction in heart-rate range was associated with sucking *per se* rather than the provision of dextrose or milk.

Sucking scores were studied in relation to length of labor and drugs or anesthetics given to the mother before delivery. The sucking scores bore no relation to these variables.

DISCUSSION

This experiment was designed to compare sucking under conditions that were thought to represent progressive approximations to a bottle feed. Was pump feeding really like a normal feed by bottle? Although the pump feed took longer and the babies gave a greater number of sucks and had to be burped more frequently, the analyses of these data showed that the differences could have arisen by chance. Eight of the 24 pump-fed babies took some milk from the bottle when it was offered at the end of the pump feed (one-eighth to one-half oz) and in this respect their performance did not differ from that of bottle-fed babies who often take more from the bottle when it is offered them by an ex-

perienced nurse at the end of a feed. Thus, despite a tendency to require a longer time, the pump feed provides a facsimile of bottle feeding.

These findings clear the way for an examination of the first hypothesis, namely: that nutritive fluids are significant determinants of sucking. The results reinforce the conclusions drawn from an earlier experiment (Dubignon and Campbell, 1968b): when nutritive fluid is delivered, marked changes in the pattern of sucking occur. It should be noticed that an analysis based upon sucking counts would show a significant drop in expression scores when dextrose was introduced on the second trial followed by a return to the non-nutritive level in the first 2 minutes of the pump feed. However, these shifts in count follow from a change in the over-all pattern of responding: the rate of sucking, for both expression and suction, fell when dextrose was introduced in the second trial and became even slower during the first 2 minutes of the feed. (In this context, rate means the rate of expression and suction responses per second during the intervals of time when the baby was actually sucking). The fall in rate was offset by an increase in time spent sucking when dextrose was introduced and a further increase during the feed. As predicted, during the feed the changes in pattern of sucking, which had already been observed to take place immediately on the addition of dextrose to the array of intraoral stimuli, were exaggerated.

The order in which the non-nutritive, dextrose and feeding conditions were presented must be considered in interpreting these results. It has been shown previously that the order of non-nutritive and dextrose trials has no effect upon the sucking patterns elicited under each condition (Dubignon and Campbell, 1968b). In the present study it was not possible to vary the order of feeding and dextrose conditions because, after a feed, it is difficult to elicit sucking reliably. In a recent study sucking for small and equal amounts of milk and dextrose has been compared when the order of trials was counterbalanced. Milk elicited higher sucking counts and more time spent sucking and both these effects were independent of the order of trials (Dubignon and Campbell, 1968c). It is therefore improbable that the order of conditions in the present experiment contributed significantly to the changes in pattern of responding.

An experiment based upon sucking counts alone would have led to the conclusion that non-nutritive sucking and sucking during a feed do not differ. One practical consequence of these results is that the study of sucking in relation to changes in stimulating conditions should be based upon measures that are more analytical than a sucking count taken over a specified length of time because a count alone does not do justice to changes in the pattern of responding.

The findings agree with those of Halverson (1923). Sameroff (1965),

who reported lower counts during non-nutritive sucking, fed his subjects for 5 minutes before the measurement of non-nutritive sucking so that the conditions were not directly comparable with the present experiment.

There were no significant changes in the ratio of expression count to suction count between the three treatments, which suggests that in both nutritive and non-nutritive conditions the mechanism of sucking is the same. However, the expression component tended to occur more frequently than the suction component. It is frequently possible to have an expression component without a matching suction component; the converse of this—suction without expression—rarely happens (cf. Sameroff, 1965). The amplitude of suction responses showed a significant increase across trials so that the babies used the greatest negative pressure during the pump feed. Expression amplitude, on the other hand, declined during feeding but not sufficiently to reach significance. One could argue, that the relative importance of suction and expression (as measured by the pressure exerted) shifts when a feed is given. However, this result should be checked using a variety of delivery systems since the pressure exerted in any given setting could conceivably be an artefact of the mechanism used to deliver milk.

Before framing any conclusions regarding the differences found between the three treatments it is instructive to re-examine the sucking rates:

	Non-nutritive	Dextrose	Feeding
Suction rate	1.55	1.26	1.16
Expression rate	1.54	1.30	1.21

The major change in rate took place between the non-nutritive and dextrose trials and a smaller drop occurred during feeding. The average consumption of fluid during the periods scored was 3.5 cc of dextrose and 11.5 cc of milk. It is possible that the small drop in sucking rate between the dextrose and feeding trials may stem from the need to swallow three times as much fluid, rather than differences in the qualities of milk and dextrose. The fact that the large difference in intake caused only a slight drop in rate makes it clear that the much larger decrease in rate between non-nutritive and dextrose trials is not likely to be due to the need to swallow 3.5 cc during the dextrose trial. Here the change in sucking pattern is clearly due to the change in intra-oral stimulation.

The changes in sucking during the feed suggest that it may be necessary to make some distinctions between the control of sucking by intra-oral stimulation and the control exerted by satiation or fatigue. It should be noted that during the pump feed the significant changes in counts and time spent sucking in the last quarter of the feed are not matched by

any change in sucking rates. Similarly, there were no changes in sucking rates before and after burping periods. It seems plausible to argue that the sucking *rate* is primarily under the control of intraoral stimulation while the *probability* of sucking is controlled by other factors as well. Thus a baby, given a feed, will react to satiation, not by sucking more slowly, but by sucking less of the time.

Both fatigue and satiation must be considered in interpreting the changes in sucking behavior that occurred at the end of the feed. Studies of non-nutritive sucking over long periods (10–15 minutes) have shown that work decrements are small and that there is complete recovery of initial sucking levels given a 1-minute rest period (Levin and Kaye, 1964, 1966). These findings are confirmed in the present study of feeding. Examination of time spent sucking and mean sucking counts for the first quarter of the feed and after the last burping period reveals that the infants showed almost complete recovery of initial sucking levels after rest.

	Suction count	Expression count	Suction time	Expression time
First quarter of feed	36.46	42.25	30.04	34.63
After last burping period	33.92	41.58	28.63	35.54

The increase in time spent sucking and sucking count after burping can be construed as recovery from fatigue. This recovery was as large in the early part of the feed as it was toward the end of the feed. On the other hand, it appears that increasing satiation is responsible for the decrease in time spent sucking and lower sucking counts at the time of the last burp as compared with the first burp. It will be remembered that during the last quarter of the feed the part of the record scored was the final 45 seconds of sucking. After a further rest the infant refused to accept the nipple. It therefore seems likely that satiation and fatigue are jointly responsible for the changes in sucking behavior that were observed at the end of the last quarter of the feed, while the refusal to accept the nipple after a further rest may be attributed to satiation alone.

The uniformity of sucking performance during the first three quarters of the feed stands in contrast to decrements in non-nutritive sucking after the consumption of amounts of fluid as small as 10–20 cc (Kaye, 1966). These effects were interpreted as due to satiation. In the present study no decrement in responding was observed until 60–90 cc of fluid had been consumed. Two interpretations of these findings are possible: either small amounts of fluid in the stomach have different effects on nutritive and non-nutritive sucking, or the decrement found by Kaye

(1966) was a contrast phenomenon, i.e., the dry nipple may be less effective in sustaining sucking after the infant has experienced the additional stimulation of fluid in the mouth.

The change in expression/suction ratio at the end of the feed indicates that satiation or fatigue induces a change in the mechanism of sucking; the suction component tends to drop out, while the expression component appears to represent chewing or mouthing of the nipple. This finding supports the conclusion that suction is the means by which bottle-fed infants usually control their food intake from standard commercial nipples as opposed to the "stripping" action that may be the effective means of getting milk from the breast.

Throughout the feed the trigger was set to fire for low-amplitude responding, so that milk was delivered for every sucking response. The mean suction and expression amplitudes recorded were in fact four times the amplitude necessary to trigger the pump. It therefore seems likely that the higher suction amplitudes recorded for the expression trigger group during the feed were a chance finding.

At first sight the pattern of changes in heart rate and heart-rate range give the impression that feeding exerts a stronger influence than non-nutritive or dextrose sucking on "state" as measured by heart rate. However, examination of the sucking scores suggests that feeding has this influence because more time is spent sucking during feeding. The plot of time spent sucking under the various conditions is virtually a mirror image of the plot of heart-rate range, and this supports the conclusion that the reduction in heart-rate variability is a function of sucking *per se* rather than the provision of nutriment.

The relation between heart rate and sucking is not quite so obvious. Nutritive sucking leads to increases in heart rate, and as sucking drops off toward the end of the feed heart rate shows a corresponding drop toward the resting level. Similar increases in heart rate during nutritive sucking have been reported in other studies, e.g., Baliassnikowa and Model (1931); Lipton *et al.* (1958). The observation that high heart rates tend to be associated with low heart-rate ranges suggests that the reduction in range may possibly be attributed to a "ceiling" effect. However, the heart rate of newborn infants can increase very considerably above the highest mean rates reported here (approximatly 147 beats per minute).

To sum up, non-nutritive sucking has been shown to differ from nutritive sucking on five of the six measures of sucking investigated in this study. In most cases, the differences were shown in both the suction and expression components of sucking. The results replicate and extend earlier findings and confirm the hypothesis that the stimuli provided by

nutritive fluids are significant determinants of sucking. The differences between non-nutritive sucking and sucking during a feed make it clear that investigations that are intended to have implications for the feeding of the newborn should be carried out in a nutritive setting. The earlier finding that non-nutritive sucking frequency is in part determined by the characteristics of the nipple cannot be generalized to infant feeding without further study.

Sucking behavior throughout a feed remained fairly uniform until the end of the last quarter when a decrease in sucking counts, time spent sucking, and suction amplitude, as well as a reduction in the frequency of the suction component of sucking marked the end of the feed. No corresponding changes in sucking rate were found, which suggests that this measure of sucking may be primarily under the control of current intra-oral stimulation.

REFERENCES

Baliassnikowa, N. J., and Model, M. M. Zur Neurologie des Saugens. *Zeitschrift fur Kinderheilkunde.*, 1931–2, **39**, 1–16. (Cited in Halverson, 1938).

Balint, M. Individual differences of behavior in early infancy and an objective method for recording them, I: Approach and the method of recording. II: Results and conclusions. *Journal of Genetic Psychology*, 1948, **73**, 57–79; 81–117.

Colley, J. R. T., and Creamer, B. Sucking and swallowing in infants. *British Medical Journal*, 1958, **2**, 422–423.

Dubignon, J. Intra-oral stimulation and sucking in the newborn. Unpublished Ph.D. Thesis: Queen's University, 1967.

Dubignon, J., and Campbell, D. Intra-oral stimulation and sucking in the newborn. *Journal of Experimental Child Psychology*, 1968 (a), **6**, 154–166.

Dubignon, J., and Campbell, D. Sucking in the newborn under three conditions: non-nutritive, nutritive and a feed. *Journal of Experimental Child Psychology*, 1968 (b), **6**, 335–350.

Dubignon, J., and Campbell, D. Discrimination of milk and dextrose by human newborns. 1968 (c) (in preparation).

Halverson, H. M. Infant sucking and tensional behavior. *Journal of Genetic Psychology*, 1938, **53**, 365–430.

Halverson, H. M. Mechanisms of early infant feeding. *Journal of Genetic Psychology*, 1944, **64**, 185–223.

Kaye, H. Effects of feeding and tonal stimulation on non-nutritive sucking in the human newborn. *Journal of Experimental Child Psychology*, 1966, **3**, 131–145.

Kaye, H. Infant sucking and its modification. In L. P. Lipsitt and C. C. Spiker (Eds.) *Advances in Child Development and Behavior*, vol. 3, Academic Press, New York: 1967.

Kron, R. E., Stein, M., Goddard, K. E., and Phoenix, M. D. Effect of nutrient upon the sucking behavior of newborn infants. *Psychosomatic Medicine*, 1967, **23**, 24–32.

Levin, G. R., and Kaye, H. Non-nutritive sucking by human neonates. *Child Development*, 1964, **35**, 749–758.

Levin, G. R., and Kaye, H. Work decrement and rest recovery during non-nutritive

sucking in the human neonate. *Journal of Experimental Child Psychology*, 1966, **3**, 146–154.

LIPSITT, L., AND KAYE, H. Change in neonatal response to optimising and non-optimising sucking stimulation. *Psychonomic Science*, 1965, **2**, 221–222.

LIPTON, E. L., RICHMOND, J. B., WEINBERGER, H. L., AND HERSHER, L. An approach to evaluation of neonate autonomic responses. *Psychosomatic Medicine*, 1958, **20**, 409.

SAMEROFF, A. J. An experimental study of the response components of sucking in the human newborn. Unpublished Ph.D. Thesis: Yale University, 1965.

TINBERGEN, N. *The study of instinct*. London and New York: Oxford Univ. Press, 1951.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 299–313 (1969)

The Effects of Time-Out from Positive Reinforcement on the Operant Behavior of Preschool Children[1]

ROBERT H. WILLOUGHBY
University of Massachusetts

Sixty preschool children participated in two experiments testing the suppressive effects of time-out from positive reinforcement (TO). In Experiment I a lever press was reinforced on a VI schedule with a 10-second TO administered after every seventh response (FR7). When the punished response was the only one available, TO depressed response rate by approximately 25%. However, this suppression was of relatively brief duration. With the introduction of an unpunished alternative, TO produced complete and lasting suppression. In Experiment II a response preference was established by reinforcing responses on one lever more frequently than those on another and then punishing responses on the preferred lever with a 10-second TO. Results showed that under punishment by TO the formerly unpreferred lever was chosen even though the frequency of reinforcement for the two levers was identical.

When positive reinforcement is withheld, the frequency of the reinforced response decreases; if an external stimulus is selectively associated with this nonreinforcement, then the response rate can be shown to decrease in the presence of that stimulus. Such a stimulus may be designated a time-out (TO) stimulus, and the length of its presentation a time-out period. In the majority of early operant investigations, TO was limited to a period separating the organism's opportunity to respond for reinforcement (Ferster and Skinner, 1957). However, more recent research has shown that TO may also be used to control the frequency of a given response by arranging it as a consequence of that response—what might be considered punishment by TO (Ferster, 1958; Baer, 1962a, 1962b; Holz, Azrin, and Ayllon, 1963; Leitenberg, 1966).

Results from the experimental laboratory (Baer, 1961, 1962a, 1962b), as well as the findings from naturalistic investigations (Wolf, Risley, and

[1]This study was part of a larger investigation submitted to the graduate school of the University of Minnesota in partial fulfillment of the Ph.D. The author wishes to express his gratitude to Willard W. Hartup for his generous advice and assistance during the conduction of this study. The research was supported in part by Predoctoral Fellowship 5-F1-MH-30, 341-02 from the National Institute of Mental Health and completed with the aid of the facilities of the Numerical Analysis Center at the University of Minnesota.

Mees, 1963; Hawkins *et al.*, 1966), have provided ample evidence that TO can be used to modify a variety of behaviors exhibited by young children. However, the methods by which TO implements such behavioral change are not entirely clear. A recent review of the literature of this topic by Leitenberg (1965) has revealed that several of the experimental findings which are frequently cited as evidence of the aversiveness of TO may also be interpreted in terms of the increased reinforcement which frequently accompanies the avoidance of TO. Thus, Leitenberg has argued, one need not conclude that TO functioned as a noxious stimulus with suppressive properties in order to explain the results from the majority of these studies; their findings are just as easily explained in terms of altered reinforcement frequency.

One method of unconfounding the variables of reinforcement frequency and TO presentation is to arrange the reinforcement schedule of the punished response such that any change in response rate does not affect the frequency of positive reinforcement per unit time. Under such conditions, any decrease in response rate could be attributed to the specific consequences of that response. If TO did function as a punishing stimulus, then under such conditions one would expect the overall rate of responding to decrease. At least one study has used this procedure with human subjects. Holz *et al.* (1963) reinforced psychiatric patients on a VI schedule while concurrently presenting a brief TO after every tenth response (FR10). Under these conditions, reinforcement frequency did not depend upon rate and the suppressive effects of TO could be evaluated independently of the frequency variable. Despite the attempt by Holz *et al.* (1963) to separate the effects of these two variables, some of their subjects did experience marked reductions in reinforcement frequency across experimental sessions. Thus, these investigators were only partially successful in their attempt to unconfound the effects of changes in reinforcement frequency from those of TO.

What may be an even more effective method of isolating the effects of the TO variable is the use of a "yoked" control procedure. In such a procedure the control subject receives the same number of reinforcements as his "yoked" experimental partner without ever receiving a discriminable period of TO. This procedure enables the investigator to separate the effects of TO from those of changes in reinforcement frequency, and permits an independent assessment of the suppressive properties of TO. The present investigation attempted to make such an assessment by studying effects of TO on the operant behavior of young children.

EXPERIMENT I

This experiment was designed to test the effectiveness of TO as a punishing event. In the conventional punishment paradigm, the punish-

ing stimulus is always response contingent and the measure of its effectiveness lies in the subsequent decline in the frequency of the antecedent response. It was the purpose of the experiment to ascertain whether TO would suppress response rate when administered contingently in a situation in which one, and later two, responses provided the means to positive reinforcement.

Method

Subjects. The *S*s were 36 children, 18 boys and 18 girls, from a middleclass, suburban preschool in the Minneapolis area.[2] Twelve *S*s were randomly assigned to each of three treatment groups: a time-out group (Group TO), a yoked-TO group (Group Y-TO), and a no-TO control group (Group C). These experimental conditions are shown in Table 1.

Apparatus. The experimental apparatus consisted of a front panel of glazed masonite 24- × 24-inches in surface area. The panel stood in an upright position and was supported by two pieces of ½-inch plywood, 12- × 24-inches connected to it at a ninety-degree angle. Extending from the front panel 4½ inches from the lower edge were two levers serving as response manipulanda. The lever on *S*'s left was made of blue laminated plastic, the lever on his right of white laminated plastic; located 15 inches above the manipulanda, and 4½ inches below the upper edge of the panel, was a row of ten stimulus lights 2 inches wide and 22 inches long. These lights were housed behind a screen of white flash glass and were lit when the apparatus was in operation.

Red poker chips were used as reinforcers. The chips were dispensed through a vertical slot located 6 inches from the left manipulandum and 8 inches from the panel's lower edge. A transparent plexiglas container served as a receptacle for the chips. Reinforcements were dispensed on a variable interval schedule with an average interval between reinforcements of 10 seconds (VI-10 seconds). The scheduling of reinforcements was controlled by a tape programming mechanism located in an adjoining room. The first response made after the reinforcement dispensing mechanism had been "primed" unlocked a holding relay and delivered a chip. Responses on the manipulanda were automatically recorded in 30-second intervals by a Presin (Moduprint) Counter. The duration of the experimental session was controlled by a predetermining counter (Sodeco, Model TCEZ4PE) which shut off the experimental apparatus automatically.

Time-out was implemented by activating a circuit which contained a stepping relay (Grason-Stadler, E3129B) and an electronic timer

[2] The author expresses his appreciation to the staff of St. Davids Preschool, Minnetonka Mills, Minnesota, Mrs. Sybil Lynch, Director for their cooperation during all phases of this research.

(Grason-Stadler, E1100H). During that part of the experiment employing TO, every seventh response operated the electronic timer for a 10-second period. During the timer's operation, reinforcement was inaccessible. The row of signal lights was extinguished and the response manipulanda were immovable. In addition, the tape programming mechanism, the Sodeco counter, and the response recording apparatus were shut off. This procedure assured that the recording periods during the TO phase of the study were equal in length to the recording periods of the control phase. At the onset of TO, the stepping relay automatically reset, causing the seventh response after TO to again operate the timer and produce another 10-second TO. Thus, in the punishment procedure, responses were reinforced on a VI-10 second schedule and punished on an FR7 schedule.

Procedure

The experiment consisted of three distinct phases, each containing three minutes of response time. The experimental conditions are presented in Table 1.

TABLE 1
DESIGN FOR EXPERIMENT I

	Phase 1	Phase 2	Phase 3
TO group			
R-1	VI-10 sec.	VI-10 sec. + 10-sec. TO	VI-10 sec. + 10-sec. TO
R-2	—	—	VI-10 sec.
Y-TO group			
R-1	VI-10 sec.	VI-10 sec. + 10-sec. "yoked" TO	VI-10 sec.
R-2	—	—	VI-10 sec.
Control group			
R-1	VI-10 sec.	VI-10 sec.	VI-10 sec.
R-2	—	—	VI-10 sec.

Phase 1: All *S*s received the same treatment in this phase. The *S* was escorted into the experimental room by *E* and seated in front of the apparatus. The *E* then showed *S* a number of small prizes and asked whether he would like to attempt to win one of them. The *E* then instructed *S* that in order to win a prize, he had to fill the plexiglas container with chips. Then *E* proceeded to demonstrate how chips would be obtained by pressing one of the response levers. The *S* was also shown that the remaining lever was immovable and could not be pressed for reinforcement. The manipulandum which was reinforced was counter-

balanced across *Ss*, with half of the *Ss* in each group receiving reinforcement on the white lever. The *S* was encouraged to respond on the manipulandum and to try to fill the container with chips. At this point, *E* retired to an adjoining room, leaving *S* alone in the experimental room. In this phase of the experiment, responses were reinforced on the VI-10-second schedule for 3 minutes. The Presin counter printed *S*'s response frequency in 30-second intervals during this baseline period.

Phase 2: In the second phase, which followed without interruption, the TO group was given a 10-second TO after every seven responses made on the manipulandum (FR7). During TO, the apparatus was inoperable and the recording equipment shut off. The VI-10-second schedule remained in effect. The introduction of the TO procedure was accompanied by a reduction in reinforcement frequency. To control for the possible effects of this reduction, a Yoked-TO condition was run. Each of the twelve *Ss* in the Y-TO condition was matched with an *S* from the TO condition and received the same number of TOs as the punished *S*. However, the Yoked-TOs had only the programming and recording equipment shut off for the 10-second TO period. For these TOs, the apparatus lights remained lit and the response lever was not locked. Thus, while the Y-TO Group received the same number of TOs as the TO Group, the TOs were not accompanied by any discriminable change in stimulus conditions. The Control Group continued to receive reinforcement on the same VI schedule and never experienced TO. The total duration of this phase, excluding TO, was 3 minutes.

Phase 3: In the third and final phase of the study, all *Ss* were provided with an unpunished response alternative. The second manipulandum was unlocked and *S* was told that he could respond on either lever for reinforcement. For the Yoked-TO and Control Groups, both manipulanda were reinforced on the VI-10-second schedule. Only the formerly locked manipulandum was unpunished in the TO condition; responses on the original manipulandum continued to be punished with a 10-second TO. After *S* had received these instructions, he was given a 3 minute period of free responding. At the end of this period, *S* was allowed to select his prize and was escorted to his classroom.

Results

Figure 1 shows the response rates for all three groups before and during the punishment procedure. All groups had a highly stable rate of response during the baseline period. Furthermore, there is no evidence that the experimental and control groups differed in the number of responses made per 30-second interval during this phase. A decrease in response rate accompanied the introduction of TO in Phase 2. Neither

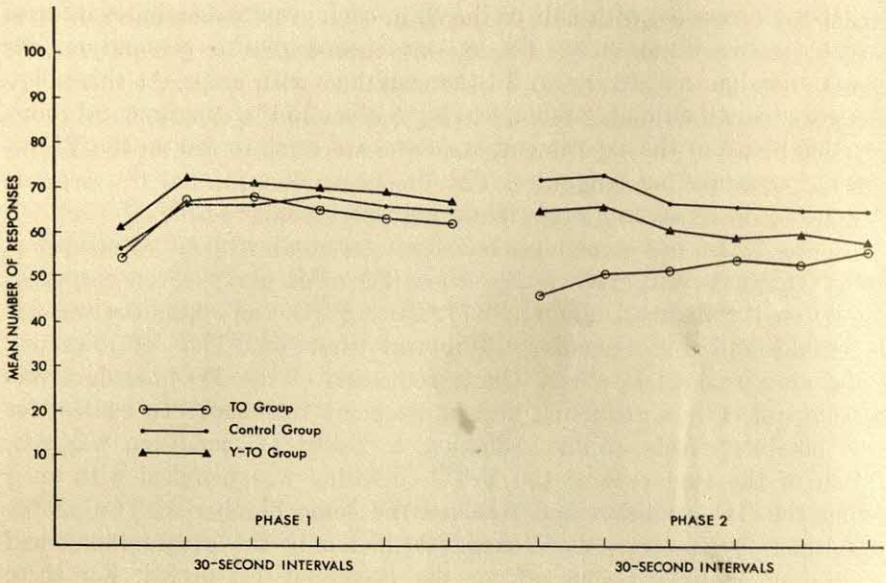


FIG. 1. Mean response rate for three groups in Phases 1 and 2 of Experiment I.

the Y-TO nor the Control Group showed any appreciable reduction in response rate during this phase.

The results of a 2 (Phases) $\times$ 3 (Groups) analysis of variance performed on the response rate data for Phase 1 and Phase 2 showed the Groups $\times$ Phases interaction to be statistically significant ($F = 6.42$, $df = 1/33$, $p < .01$). Subsequent analyses of this interaction for simple-effects followed procedures outlined in Winer (1962). These analyses revealed that only the TO Group differed significantly in response rate from Phase 1 to Phase 2 ($F = 13.85$, $df = 1/33$, $p < .01$). An additional between-groups analysis of the interaction showed that three groups did not differ significantly in rate of response during the baseline period ($F = .701$, $df = 2/66$), but that the difference in rate was statistically significant during the punishment period ($F = 5.82$, $df = 2/66$, $p < .01$). Individual comparisons between groups in Phase 2 showed that the Y-TO Group and the Control Group were not significantly different in the total number of responses made during this phase ($F = 1.47$, $df = 1/66$) but that both of these groups had a significantly higher response frequency during the punishment phase than the TO group (TO vs. Y-TO, $F = 4.66$, $df = 1/66$, $p < .05$; TO vs. C, $F = 11.35$, $df = 1/66$, $p < .01$).

To ascertain which of the six response measures in Phase 2 were significantly below the Phase 1 rate of the TO Group, all twelve

measures of response rate were compared by the Newman-Kuels procedure (Winer, 1962). This analysis revealed that the response rate during the first thirty seconds of Phase 2 was significantly below the response rate of each of the six recording periods in Phase 1 ($p < .01$, $df = 12/120$). However, the response rates during the remainder of Phase 2 were not significantly below the baseline rate.

The response rates for the three groups during Phase 3 on both the punished and unpunished manipulanda are presented in Figure 2. As

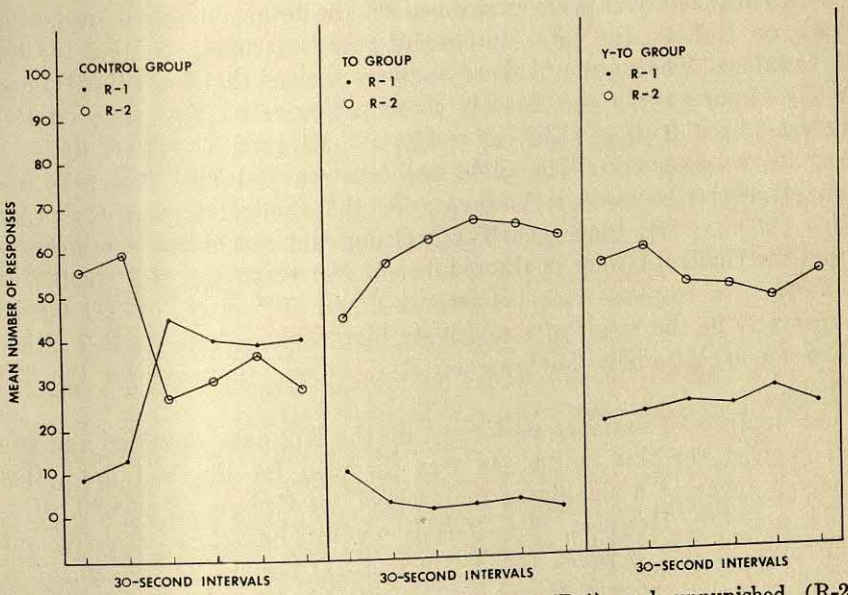


FIG. 2. Mean response rate on the punished (R-1) and unpunished (R-2) manipulanda in Phase 3 of Experiment I.

the figure illustrates, the rate of the punished response (R-1) was approximately zero for the TO Group throughout the entire third phase. This is not the case, however, for the Y-TO and Control Groups which received no TO for R-1.

An analysis of variance performed on the rate data for R-1 during Phase 3 revealed a significant main effect across Groups ($F = 35.88$, $df = 2/33$, $p < .01$). Subsequent comparisons between the three treatment groups revealed that the Y-TO Group did not differ significantly from the Control Group in R-1 rate ($F = 1.43$, $df = 1/198$), but that the TO Group had a significantly lower R-1 rate than either the Y-TO Group ($F = 56.67$, $df = 1/198$, $p < .01$) or the Control Group ($F = 76.05$, $df = 1/198$, $p < .01$).

Figure 2 shows that the R-1 rate was low for all three treatment

groups immediately after the new manipulandum (R-2) was introduced. The preference for R-2 by all groups during the early part of Phase 3 resulted in a significant Groups × Intervals interaction during this phase ($F = 3.89$, $df = 10/165$, $p < .01$). The preference for R-2 shown by the Control *Ss* in the early part of Phase 3, because of its rather short duration, would appear to represent a novelty effect. In contrast, the strong preference shown for R-2 by the TO Group remained throughout the entire phase.

Statistical analyses were performed on the total number of responses made on R-1 vs. the total number of responses made on R-2 for the three groups. The results of these analyses showed that both the TO and Y-TO Groups made significantly more responses on R-2 than on R-1 during Phase 3 ($t = 11.96$, $df = 11$, $p < .01$, and $t = 3.47$, $df = 11$, $p < .01$, respectively). The difference between R-1 and R-2 rate was not stastistically reliable, however, for the Control Group ($t = 1.93$, $df = 11$, $p > .05$). Since the Y-TO Group did not differ significantly from the Control Group in the number of R-1 responses, this significant difference in response rate between R-1 and R-2 in the Y-TO Group appears to be the result of a relatively high response rate on R-2 rather than an exceptionally low rate on R-1, as was the case for the TO Group.

An analysis of variance performed on the R-2 data provided statistical support for this notion. As was the case for the R-1 data, this analysis revealed a significant main effect for Groups ($F = 4.20$, $df = 2/33$, $p < .05$). However, this effect is not attributable to a significant difference in rate between the Y-TO and TO Groups ($F = 2.36$, $df = 1/198$, $p > .05$) as was found for R-1. Rather, the significant Groups effect is the product of a low response rate on R-2 by the Control *Ss* relative to the R-2 rate of the TO *Ss* ($F = 9.46$, $df = 1/198$, $p < .01$). This significant preference for R-2 shown by the Y-TO Group, despite the equal reinforcement frequencies of the two responses, indicates that the reduced reinforcement on R-1 in the second phase of the study had some influence upon the response preference of the Y-TO *Ss* in the subsequent phase.

The overall response rates for in Phases 1 and 3 were compared for the three treatment groups. Only the Y-TO Group showed any substantial increase in response rate during Phase 3 above the original baseline rate. This difference in rate from Phase 1 to Phase 3 was found to be significant for the Y-TO Group ($t = 2.78$, $df = 11$, $p < .02$) and suggests the presence of a positive "contrast" effect (Reynolds, 1961). The response rates in the third phase were not significantly different from baseline for either the TO Group ($t = 1.69$, $df = 11$, $p > .05$) or Control Group ($t = 1.73$, $df = 11$, $p > .05$).

EXPERIMENT II

The results of Experiment I clearly demonstrate that punishment by TO produced a high degree of response suppression when *S* was provided with an unpunished response alternative. However, the possibility still remains that the preference for the manipulandum not associated with TO during Phase 3 resulted from the slightly higher frequency of reinforcement which occurred on that manipulandum.

In order to provide a totally unambiguous interpretation of these findings, a second experiment was performed. In this study, the reinforcement frequencies of the punished and the unpunished responses were equated. This method of control was accomplished by using a VI-20-second reinforcement schedule for the unpunished response and a VI-10-second reinforcement schedule for the punished response; TOs for the latter response were equal to 10 seconds in duration. Thus, both the punished and unpunished responses had the same reinforcement frequency but one had an accompanying TO period.

Method

Subjects. Twenty-four children, 12 boys and 12 girls, from the same suburban preschool sample served as *S*s. Half of the *S*s were randomly assigned to a Time-out (TO) condition and the remaining *S*s were assigned to a no-TO, Control, condition.

Apparatus. Major changes in the experimental apparatus involved only the reinforcement programming equipment. An additional (VI-20 second) reinforcement schedule was programmed concurrently with the original (VI-10-second) reinforcement schedule. Responses made on one of the manipulanda were reinforced on the VI-10-second schedule; the other manipulandum was reinforced on the VI-20-second schedule.

In addition to these changes, the schedule of TO administration was also modified. To equate the two responses for reinforcement frequency, a response contingent TO of 10-second duration was programmed on the VI-10-second manipulandum on a VI-10-second schedule. Thus, the number of TOs was equal to the number of reinforcements dispensed on that manipulandum within the session. The intervals between TO presentation were arranged so that their occurrence was not temporarily associated with the presentation of reinforcement.

Procedure

The study was conducted in four phases (see Table 2). Phase 1 served as a training period and was 2 minutes in length. During this period, *S* was given one minute of response time on each manipulandum. At the end of this minute, *E* counted the chips received and placed them

TABLE 2
DESIGN FOR EXPERIMENT II

	Phase 1	Phase 2	Phase 3	Phase 4
TO group				
R-1	Training VI-10 sec.	VI-10 sec.	Training VI-10 sec. + 10-sec. TO	VI-10 sec. + 10-sec. TO
R-2	Training VI-20 sec.	VI-20 sec.	Training VI-20 sec.	VI-20 sec.
Control group				
R-1	Training VI-10 sec.	VI-10 sec.	Training VI-10 sec.	VI-10 sec.
R-2	Training VI-20 sec.	VI-20 sec.	Training VI-20 sec.	VI-20 sec.

in a pile beneath the appropriate manipulandum. Then *S* responded for one minute on the remaining manipulandum and this counting procedure was repeated. After this training was completed, *S* was instructed that he could respond on either manipulandum for reinforcement. These instructions initiated Phase 2 of the experiment—a response preference period. This period lasted three minutes during which *S*'s response preferences were recorded in 30-second intervals. The *S*s in both groups were given the same procedures during Phases 1 and 2 of this study.

In Phase 3, the experimental procedures were different for the two groups. Phase 3 consisted of another 2-minute training period. In this period, the TO Group received a 10-second TO for each reinforcement on the VI-10-second schedule. Responses on the other manipulandum remained reinforced on the VI-20-second schedule. The procedure employed with the Control Group in this second training period was identical to that used in the initial training session.

The training period was followed by another 3-minute preference period (Phase 4), in which responses made on one manipulandum were reinforced on the VI-20-second schedule, and the responses made on the other manipulandum were reinforced on the VI-10-second schedule plus a 10-second TO per reinforcement. The end of Phase 4 marked the termination of the experimental session.

Results

Figure 3 shows the response rates on both manipulanda for the TO and Control Groups during Phases 2 and 4. Before punishment training, the more frequently reinforced manipulandum was overwhelmingly

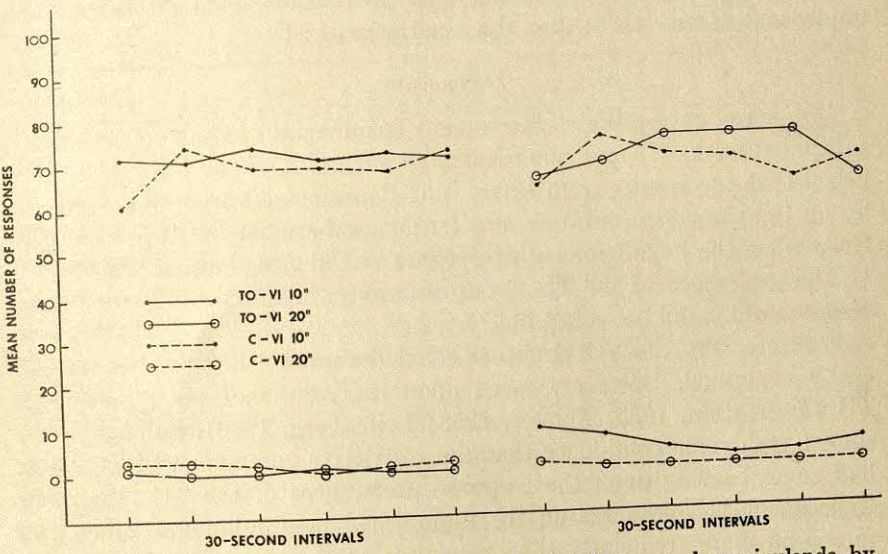


FIG. 3. Mean response rate on VI-10-second and VI-20-second manipulanda by TO and Control Groups in Phases 2 and 4 of Experiment II.

preferred by both groups. However, when this preferred response was punished with a 10-second TO, its rate declined to near zero.

The results of statistical analyses performed on the rate data from the VI-10-second manipulandum showed that the Groups × Phases interaction was significant ($F = 190.53$, $df = 1/22$, $p < .001$). A further analysis of this interaction revealed that the TO and Control Groups did not differ in rate of response prior to punishment training ($F = 1.67$, $df = 1/22$, $p > .05$), but were significantly different in rate after TO had been introduced ($F = 182.11$, $df = 1/22$, $p < .001$). The procedures employed in this experiment preclude the possibility that these rate differences during Phase 4 could have resulted from an imbalance in reinforcement frequency. Such differences are better explained in terms of the avoidance of TO.

Phase 4 was characterized by a reduction in reinforcement frequency for the TO Group. This reduction was coincident with elevations in response rate for several *S*s in the TO condition, and was frequently accompanied by verbalizations indicative of a heightened motivational state (e.g., "Hurry up, Chip. C'mon Chip,"). It is interesting that while several *S*s showed quite substantial increments in response rate when the reinforcement schedule became less favorable, they rarely made a single response on the manipulandum which produced TO. These

findings suggest that while a reduction in reinforcement frequency was unpleasant, it was less so than the experience of TO.

Discussion

Through equating the reinforcement frequencies of the responses both associated with TO and not associated with TO, Experiment II established that the results from Phase 3 of Experiment I were not produced by an imbalance in reinforcement frequency between the two responses. Even when the payoff for both responses was identical, as in Experiment II, the preference of the TO Group was overwhelmingly in favor of the response which did not result in TO.

Precisely why the TO Group selected the unpunished response is not easily explained. However, based upon the results of other studies of TO (Herrnstein, 1955; Ferster, 1958; Leitenberg, 1966), it may be assumed with some confidence that the restrictive component of TO itself had some bearing upon the response preferences of the TO Group. In Experiment II, responses on the unpunished manipulandum were reinforced no more frequently than responses on the manipulandum associated with TO. However, these responses were not subject to frequent interruption as were those associated with TO. In a more extensive investigation of TO with young children (Willoughby, 1967), it has been shown that the frequent occurrence of TO for responses on one manipulandum will elicit high rate response bursts on a second unreinforced manipulandum. Since responses on this latter manipulandum have absolutely no extrinsic reinforcement value, whereas responses on the former manipulandum were reinforced on a highly favorable schedule, it would seem reasonable to conclude that such response bursts are indicative of some form of frustrative reaction to TO itself. Such results are not unlike those reported by Adelman and Maatch (1955) and Wagner (1959). In these studies it was shown that rats would elicit escape responses to TO, and stimuli associated with TO, in the absence of any extrinsic reinforcement for such responses. If TO can elicit such "escape" responses without any extrinsic reinforcement, as these findings indicate, then the reinforcement of responses serving as alternatives to TO should only serve to enhance the occurrence of such "escape" behaviors.

The results of the first experiment clearly demonstrate that the suppressive capacity of TO is largely determined by the availability of a positive response alternative. When the punished response provided the only means of reinforcement, the rate of that response was reduced by approximately 25% of its original baseline rate during the earliest part of the punishment phase. However, with the introduction of an

unpunished response alternative, this response was virtually eliminated. These findings are consistent with those from studies which have used other stimuli such as noise (Herman and Azrin, 1964; Azrin and Ayllon, 1966) as punishment, and imply that both TO and noise function as relatively ineffective punishments in the single response situation.

Results such as those reported for Experiment I raise the question as to why TO is not a very effective punishment in the single response situation. Merely stating that TO is not "aversive" does little in the way of explaining why other stimuli, electric shock for example, are highly effective in this regard while TO is not. Apparently, stimuli like shock evoke behaviors which are somewhat different from those elicited by less aversive events as TO. One such difference between the two stimuli may be their relative tendencies to elicit responses which are incompatible with the responses being punished. There is some question as to whether TO can elicit such incompatible responses. Rather than serving to elicit the withdrawal behaviors characteristic of noxious physical stimuli, TO has been shown to cause an increase in their rate (Herrnstein, 1955; Leitenberg, 1966).

While it is true that the punished response showed no increase in rate from Phase 1 to Phase 2 in Experiment I, the repeated presentation of TO in the second phase was accompanied by an increase in the rate of that response. Despite the fact that this rate increase lead to the greater incidence of TO and, consequently, to a greater loss of reinforcement per unit time, the rate of the punished response rose steadily during the second phase. By the end of Phase 2 the rate of the punished response approached its original baseline rate. This finding is somewhat perplexing, since a rate increase under an FR schedule also increases the number of TOs received. When *S* has some control over TO frequency, as he does with an FR schedule, it certainly is beneficial to keep the occurrence of nonreinforcement to a minimum. Yet this was not observed to be the case in Experiment I. It seems quite likely that the heightened motivational level engendered by the repeated experience of TO resulted in an increased tendency to elicit higher rates of response and, consequently, to increase the occurrence of nonreinforcement. Thus, it may well be that TO elicits responses which compliment rather than compete with those responses undergoing punishment, and that this tendency toward heightened responding increases with repeated presentations of TO.

The practical considerations to be derived from this study are clear; if TO is used as a method of eliminating undesired behaviors in children, then the child should also be given a second unpunished behavior in addition to making TO a consequence of the undesired response. If

this is not the case, and the punished response provides the only available means of reinforcement, it is unlikely that TO will have lasting suppressive effects upon the occurrence of the undesired behavior.

REFERENCES

ADELMAN, H. M., AND MAATCH, J. L. Resistance to extinction as a function of the type of response elicited by frustration. *Journal of Experimental Psychology,* 1955, **50,** 61–65.

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situation. *Psychological Bulletin,* 1958, **55,** 102–119.

AYLLON, T., AND AZRIN, N. H. Punishment as a discriminative stimulus and conditioned reinforcer with humans. *Journal of Experimental Analysis of Behavior,* 1966, **9,** 411–419.

BAER, D. M. Effect of withdrawal of a positive reinforcement on an extinguishing response in young children. *Child Development,* 1961, **32,** 67–74.

BAER, D. M. A technique for the study of social reinforcement in young children: Behavior avoiding reinforcement withdrawal. *Child Development,* 1962a, 33, 847–858.

BAER, D. M. Laboratory control of thumbsucking by withdrawal and representation of reinforcement. *Journal of the Experimental Analysis of Behavior,* 1962b, **5,** 525–528.

FERSTER, C. B. Control of behavior in chimpanzees and pigeons by time-out from positive reinforcement. *Psychol. Monographs,* 1958, **72,** No. 8 (Whole No. 461).

FERSTER, C. B., AND SKINNER, B. F. *Schedules of Reinforcement.* Appleton, New York, 1957.

HAWKINS, R. P., PETERSON, R. F., SCHMEID, EDNA, AND BIJOU, S. W. Behavior therapy in the home: Amelioration of problem parent-child relations with the parent in a therapeutic role. *Journal of Experimental Child Psychology,* 1966, **4,** 99–107.

HERMAN, R. L., AND AZRIN, N. H. Punishment by noise in an alternate response situation. *Journal of the Experimental Analysis of Behavior,* 1964, **7,** 185–188.

HERRNSTEIN, R. J. Behavior consequences of the removal of a discriminative stimulus associated with variables interval reinforcement. Unpublished doctoral dissertation, Harvard University, 1955.

HOLZ, W. C., AZRIN, N. H., AND AYLLON, T. Elimination of behavior of mental patients by response-produced extinction. *Journal of the Experimental Analysis of Behavior,* 1963, **6,** 407–412.

LEITENBERG, H. Is time-out from positive reinforcement an aversive event? A review of experimental evidence. *Psychological Bulletin,* 1965, **64,** 428–441.

LEITENBERG, H. Conditioned acceleration and conditioned suppression in the pigeon. *Journal of the Experimental Analysis of Behavior,* 1966, **9,** 205–212.

REYNOLDS, G. S. Behavioral contrast. *Journal of the Experimental Analysis of Behavior,* 1961, **4,** 57–71.

WAGNER, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *Journal of Experimental Psychology,* 1959, **57,** 130–136.

WILLOUGHBY, R. H. The effects of time-out from positive reinforcement on the operant behavior of preschool children. Unpublished doctoral dissertation, University of Minnesota, 1967.

WINER, B. J. Statistical Principles in Experimental Design. McGraw-Hill, New York, 1962.

WOLF, M., RISLEY, T., AND MEES, H. Application of operant conditioning procedures to the behaviour problems of an autistic child. *Behavioral Research and Therapy*, 1964, **1**, 305–312.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 314–325 (1969)

Discrimination Learning in Children as a Function of Reinforcement Condition, Task Complexity, and Chronological Age[1,2]

GROVER J. WHITEHURST
University of Illinois, Urbana, Illinois

In a two-choice paired-presentation discrimination task, 120 children at two age levels were either reinforced for correct responses, punished for incorrect responses, or rewarded for correct responses and punished for incorrect responses at either of two levels of task complexity. Reinforcement and punishment were nonverbal and were accomplished by either presenting or removing tokens. Task complexity was varied in terms of the number of irrelevant stimulus dimensions to which a response could be made. Highly significant treatment effects favored the combination of reinforcement and punishment, with no difference appearing between the reinforcement group and the punishment group. This relationship was constant across both age and task complexity, with no significant interactions occurring.

A large number of studies using animals, adults, and children as subjects (*S*s) have been concerned with the question of the relative effectiveness of various combinations of positive reinforcement and punishment in discrimination tasks. In the prototype experiment, the *S* is forced to choose between two or more alternatives over a series of trials. One group of *S*s receives a positive reinforcer contingent upon choosing the alternative that has been defined as correct and receives some neutral stimulus consequence following an incorrect response (condition R). Another group of *S*s receives a punishing stimulus contingent upon choosing the alternative defined as incorrect and receives a neutral stimulus consequence following a correct response (condition P). A third group of *S*s receives both a positive reinforcer for a correct response and a punishing stimulus for an incorrect response (condition R-P). All three groups of *S*s are run until they reach some criterion of accuracy and then the efficiency of acquisition is compared.

[1] This research was conducted while the author was a USPHS Trainee in Child Psychology, grant number MH-8298 from the National Institute of Mental Health.
[2] This paper is based on a dissertation submitted to the University of Illinois, Urbana in partial fulfillment of the requirements for the M.A. degree. The writer expresses his appreciation and thanks to Drs. Sidney Bijou and Morton Weir for their assistance and advice throughout the present study.

The results from the animal literature (Hoge and Stocking, 1912; Warden and Aylesworth, 1927; Wischner, Fowler, and Kushnick, 1963) clearly indicate that condition R-P leads to faster acquisition of a discrimination than either condition R or condition P. The studies that have used children as *Ss* (Brackbill and O'Hara, 1958; Curry, 1960; Meyer and Offenbach, 1962; Meyer and Seidman, 1960, 1961; Penney and Lupton, 1961; Spence, 1966; Spence and Segner, 1967; Stevenson, Weir, and Zigler, 1959; Terrell and Kennedy, 1957) are more ambiguous.

In the preponderance of studies that have used children as *Ss*, the finding has been that the R condition leads to slower acquisition of a discrimination than either the P condition or the R-P condition. There has been no agreement on the relative effectiveness of the P condition and the R-P condition. As a result of weaknesses in all these studies, the question of the relative effectiveness of different reinforcement combinations remains unanswered.

The following points have proved to be important sources of ambiguity in interpreting previous studies. (1) In studies that have used verbal reinforcers and punishers (Curry, 1960; Meyer and Offenbach, 1962; Meyer and Seidman, 1960, 1961; Spence, 1966), the words "right" and "wrong" have been assumed to have reinforcing and aversive properties, respectively. In view of the complex and unknown reinforcement history that each child has with these words, such a procedural assumption appears unwarranted. Conclusions, based on these verbal studies, about reinforcement and punishment as processes in children's discrimination learning seem tenous. (2) In studies that have used nonverbal reinforcement and punishment (Brackbill and O'Hara, 1958; Penney and Lupton, 1961; Spence and Segner, 1967; Stevenson *et al.*, 1959), the procedures for presenting punishing stimuli have differed from the procedures for presenting reinforcing stimuli, such that the different reinforcement and punishment conditions may have differed in their potential for distracting the *Ss* from the experimental stimuli. This distractibility variable could account for the results in these nonverbal studies. (3) Instructions to *Ss* have varied in their completeness in previous studies, and may have interacted with the reinforcement parameters to determine the obtained differences. This sort of interaction has been demonstrated by Spence and Segner (1967). (4) Some studies have used tasks of unspecifiable complexity when task complexity has been shown to be an important variable determining the magnitude of the differences obtained among reinforcement conditions in discrimination tasks (Meyer and Offenbach, 1962).

The present study was designed to correct some of the weaknesses in previous studies in order to obtain a more definitive answer to the question of the relative effectiveness in discrimination learning of reinforcement, pun-

ishment, and reinforcement in combination with punishment. Nonverbal reinforcers and punishers were chosen to rule out, as much as possible, the effects of prior reinforcement history. A token system was instituted as a method of reinforcement. The token system allowed the use of back-up reinforcers that were considered more reinforcing to the *Ss* than the candy or praise used in previous studies. Response cost was used as the method of punishment. Punishment was the loss of a token while reward was the gain of a token. This method allowed procedurally equivalent reinforcement and punishment, which in turn removed any possibility for distraction being differentially associated with the mode of presentation of either reward or punishment. (It should be noted that punishment in terms of a removal operation may not be functionally equivalent to punishment in terms of the presentation of some aversive stimulus, and the results of this experiment must be interpreted in that light.) The complexity of the experimental task was made specifiable in terms of the number of possible stimulus dimensions to which a response could be made. In consideration of the findings of Spence and Segner (1967), the preliminary instructions fully explained the response contingencies to the *Ss*.

Two additional variables besides reinforcement condition included in the study were age and complexity. Complexity was included as a two-level variable because of its probable importance in determining the magnitude of the differences obtained (Meyer and Offenbach, 1962). Age was included as a two-level variable to ensure that the results obtained would not be age-specific, and to partially assess any developmental trends that might occur in the magnitude and direction of the differences obtained.

Thus, the study was designed to utilize a three-way analysis of variance. There were three levels of the treatment factor, two levels of complexity, and two levels of age.

The only prediction made was that the R-P condition should lead to more rapid acquisition of the discrimination than either the R condition or the P condition. It was reasoned that since incorrect responding was to be punished in the R-P condition, it should be eliminated more quickly than in the R condition in which incorrect responding was to be extinguished. Punishment has been shown to be a more effective method of eliminating behavior than extinction (Holz, Azrin, and Ayllon, 1963). An analysis of the P condition suggested that if responding to the correct stimulus was to be strengthened, it would be through the action of subtle conditioned reinforcers such as successful task completion and compliance with the experimenter's instructions. It was reasoned that the reinforcing strength of these conditioned reinforcers would be weak relative to the material reinforcers present in the R-P condition. Therefore, acquisition in the P condition was predicted to be slower than in the R-P condition.

METHOD

Subjects

A total of one hundred and twenty subjects were used in the experiment. This total was evenly divided into sixty five- and six-year-olds and sixty eight- and nine-year-olds. There were sixty-three males in the sample and fifty-seven females. All the *Ss* were drawn from the same grammar school located in a middleclass neighborhood.[3]

The *Ss* came individually to a trailer located a few yards from their school. The experiment was conducted in one room of the trailer with dimensions of approximately 8 feet × 10 feet. There were two tables and two chairs within the room. One table was located against the wall and reinforcers were arranged on it in three distinct piles. In the pile on the left ⅓ of the table were located reinforcers of about ten cents in value. Among the ten-cent reinforcers were baby dolls, packages of marbles, cooking utensils, watches, tops, police badges, balloons, modeling clay, and crayons. In the pile on the middle ⅓ of the table were located reinforcers of about five cents in value. Among these five-cent toys were rings, keychains, and plastic jewelry. In the pile on the right ⅓ of the table were reinforcers of approximately one cent in value. The one-cent toys were small plastic toys such as baseballs, trucks, and kitchen utensils. The second table was located in the center of the room. The principal apparatus for the experiment was on this table. On one side of the table was a chair for the experimenter (*E*) and on the other side was a chair for the *Ss*.

The apparatus on the table consisted of a sliding tray for presenting stimuli and a plastic dish into which tokens could be dispensed; both were built into an upright plywood screen equipped with a one-way mirror. The plywood screen was painted black and was approximately 3 feet high and 3½ feet wide; it served to shield the *E* from the *S*'s view. The sliding tray was painted white and was located at the base, and in the middle of the screen. It was so constructed that it could be withdrawn completely form the *S*'s side of the apparatus, baited with a pair of stimuli, and then pushed back out, revealing the stimuli to the *S*. On the screen to the *S*'s right, and just to the right and slightly above the aperture for the tray, was located a clear plastic dish into which tokens could be dispensed or from which tokens could be removed. An aperture in the plywood screen was located just above the plastic dish. The *E*'s hand appeared through this aperture to either dispense or remove a token from the *S*'s dish. Black cloth, attached over the aperture on the

[3] The writer thanks the principal, teachers, and students of Kenwood School in Champaign, Illinois for their generous co-operation and assistance which made this study possible.

S's side, prevented the *S* from seeing anything except the *E*'s hand during dispensation or removal of tokens. Centered in the middle of the screen and just above the sliding tray was a one-way mirror that allowed a clear view of the *S* and the extended stimulus tray.

Tokens used in the experiment were colored wooden beads, slightly smaller than marbles.

Sixteen pairs of experimental stimuli were used. They were constructed from one-quarter inch plywood. Each pair consisted of one large stimulus (an area of eight square inches) and one small stimulus (an area of four square inches). In addition, each stimulus within a pair was a given color (red or blue) and a given shape (circle or square).

Procedure

Ten *S*s within each of the two age groups were randomly assigned within ages to each of six experimental groups comprised of three treatment levels and two levels of task complexity.

Under the positive reinforcement treatment condition (R), a correct response was followed immediately by *E* dispensing a token into the *S*'s dish and next by withdrawal of the stimulus tray and its re-presentation for a new trial; an incorrect response was followed only by withdrawal of the tray and its re-presentation for a new trial. Under the positive reinforcement-punishment treatment condition (R-P), a correct response was treated as in the R condition and an incorrect response was followed immediately by *E* removing a token from the *S*'s dish and next by withdrawal of the stimulus tray and its re-presentation for a new trial. Under the punishment treatment condition (P), a correct response was followed only by withdrawal of the stimulus tray and its re-presentation for a new trial; an incorrect response was treated as in the R-P condition. Intertrial intervals for all three treatment conditions were approximately four seconds in duration. The correct response for all experimental groups was defined as the *S* placing his finger or hand on the larger of two stimuli presented on each trial.

In the first level of task complexity (simple), with one irrelevant dimension, only one pair of stimuli was used across trials for any given *S*. Color (one half the *S*s had red, one half blue) and shape (one half the *S*s had a circle, one half a square) were constant across trials with size being the relevant dimension and position (left or right) being the one irrelevant dimension. Position of the correct stimulus on any trial was determined by means of a Gellermann series (Gellermann, 1933), and was the same for each *S*. In the second level of task complexity (complex), with three irrelevant dimensions, all sixteen pairs of stimuli described in the apparatus section were used for any given *S*. Size, as before was the rele-

vant dimension, with position, color, and shape being the three irrelevant dimensions. The series of pairs of stimuli presented on succeeding trials was decided randomly, and was the same for each *S*. The positions of the correct stimulus on each trial matched those at the simple complexity level.

At the younger age level, kindergarten children were used. At the older age level, third-grade children were used.

Each child was given trials on the two choice discrimination problem until he reached the criterion of ten consecutive trials with correct responses or until he reached a total of sixty trials, whichever came first. This occurred in one session of between five and fifteen minutes in length. The data from all *S*s who made no errors was dropped from the study since differential reinforcement contingencies could not be applied when no errors were made. The total *N* of one hundred and twenty does not reflect those thirty *S*s who were dropped.

The children came individually to be tested. Each child was seated at the apparatus and the *E* read the appropriate instructions. Parts of these instructions were identical for all treatment groups, and parts differed. The instructions follow, with the differing parts indicated in parentheses and attributed to the relevant treatment condition:

"We're going to play a game and I'll show you how to play. I'll be back here and I'll push this tray out with two objects on it. Each time I push the tray out I want you to put your finger on the one that you think is correct. When you choose the right one (Group P . . . I won't do anything) (Group R and R-P . . . my hand will come out of here and I'll drop a bead in your dish, like this). When you choose the wrong one (Group R . . . I won't do anything) (Group P and R-P . . . my hand will come out of here and I'll take a bead out of your dish, like this). Do you understand? Each time you'll choose one by putting your finger on it. When you're right (Group P . . . I won't do anything) (Group R and R-P . . . I'll give you a bead). When you're wrong (Group R . . . I won't do anything) (Group P and R-P . . . I'll take one of your beads away)."

"When you're through playing you can trade in the beads you've won for one of these toys over here. Come on over and I'll show you what I've got. In this pile are my nicest toys; you have to win a lot of beads to get one of these. In this pile are some toys that aren't quite as nice and you don't have to win quite as many beads to get one of these. In this pile are my worst toys; you only have to win a few beads to get one of these. So, the idea of the game is to play just as well as you can so you will have enough beads to get one of the nicest toys."

"You can sit back down now. Remember each time put your finger on

the one you think is correct. When you're right (Group P . . . I won't do anything) (Group R and R-P . . . I'll give you a bead). When you're wrong (Group R . . . I won't do anything) (Group P and R-P . . . I'll take one of your beads away)."

After the child completed his game, regardless of the number of beads in his dish, the *E* said, "You did very well. You have just enough beads to get one of the nicest toys. Pick one."

Each child started the experiment with thirty tokens already in his dish. All questions were answered with the relevant part of the instructions.

RESULTS

The response measure used in the analysis was trials through a criterion of ten successive correct responses, which varied between a possible lower limit of eleven and a possible upper limit of sixty. A total of thirty *S*s were dropped because they had errorless scores of ten; these *S*s were approximately evenly distributed among the treatment conditions; seventy-five per cent of these errorless scores were obtained on the simple discrimination.

As indicated by Tables 1 and 2, all the main effects of the experiment were highly significant. Over levels of complexity and age, the R-P condition led to significantly faster acquisition than either the R condition or the P condition, with these last two conditions resulting in non-significantly different trials to criterion. Over reinforcement conditions and levels of complexity, the older *S*s acquired the discrimination significantly faster than the younger *S*s. Over reinforcement conditions and age, the discrimination with only one irrelevant dimension was acquired significantly faster than the discrimination with three irrelevant dimensions.

TABLE 1

MEAN TRIALS THROUGH CRITERION FOR EACH EXPERIMENTAL GROUP, TREATMENT LEVEL, AND COMPLEXITY LEVEL

	Reward mean	Punish mean	Reward-Punish mean	Complexity mean
Simple:				22.4
Kindergarten	38.7	29.6	17.0	
Third grade	17.5	17.8	13.9	
Complex:				30.1
Kindergarten	40.4	38.9	16.1	
Third grade	29.2	36.0	20.2	
Treatment Mean	31.6	30.6	16.8	

TABLE 2
SUMMARY OF ANALYSIS OF VARIANCE APPLIED TO MEAN TRIALS THROUGH CRITERION

Source	*df*	*MS*	*F*
Reinforcement condition	2	2700.93	11.73[b]
Complexity	1	1786.41	7.76[a]
Age	1	1771.01	7.69[a]
Reinforcement × Complexity	2	313.01	1.36
Reinforcement × Age	2	698.06	3.03
Complexity × Age	1	567.67	2.47
Reinforcement × Complexity × Age	2	4.98	<1
Within treatments	108	230.29	
Total	119		

[a] p less than .01.
[b] p less than .001.

None of the interactions was significant at the .05 level of confidence. However, the reinforcement by age interaction reached the .06 level. An examination of Table 1 shows that the relative positions of the R and P groups were reversed between ages. For the kindergarten *S*s, the P condition led to somewhat faster acquisition than the R condition (means of 34.3 and 39.6, respectively), while for the third grade *S*s, the R condition led to faster acquisition than the P condition (means of 23.4 and 26.9, respectively). The raw data shows that these results were caused by an abnormally high number of younger *S*s, fifty per cent, who did not achieve criterion within the allotted sixty trials in the R condition.

An additional analysis of variance was performed on the mean number of correct responses over blocks of ten trials each. It was hypothesized that the sensitivity of such an analysis to the course of learning might reveal significant interactions that failed to emerge in the analysis based on trials to criterion. Such was not the case. The significant *F*s were those for reinforcement condition, complexity, age, and trials. None of the *F*s for interactions with blocks of trials approached significance.

DISCUSSION

The results of the present study are in agreement with the findings of Brackbill and O'Hara (1958) and Stevenson *et al.* (1959). Both these studies used children as *S*s, both used nonverbal reinforcement and punishment, and both used response cost as the punishment procedure. However, neither of these studies included a P only condition. The results are only in partial agreement with the remainder of the child literature on this topic (Curry, 1960; Meyer and Offenbach, 1962; Meyer and Seidman, 1960, 1961; Penney and Lupton, 1961; Spence, 1966; Spence and Segner, 1967). More specifically, the principal area of agreement with all

the earlier findings is that the R-P condition leads to faster acquisition of a discrimination task than does the R condition.

The most important lack of correspondence with earlier child studies concerns the relative effectiveness of the P condition. One of the nonverbal studies (Spence and Segner, 1967), and most of the verbal studies (Curry, 1960; Meyer and Offenbach, 1962; Meyer and Seidman, 1960, 1961; Spence, 1966) have found no significant differences between the R-P and the P conditions. These same studies have been unanimous in the finding that the P condition leads to faster acquisition than the R condition. Both these results (R-P equal to P; P superior to R) are in disagreement with the present findings. Specifically, the results from this experiment indicate that the R-P condition leads to faster acquisition than the P condition, and that the P condition yields a trials to criterion mean that is not significantly different from the mean of the R condition. One nonverbal study (Penney and Lupton, 1961) found the P condition to lead to faster acquisition than the R-P condition. This finding is also inconsistent with the present data.

A critique of the studies using verbal reinforcement parameters has already been made (see introduction), and no more will be said here except that they are probably not applicable to generalizations about the processes of reinforcement and punishment.

The two nonverbal studies that provide findings disparate with the present study deserve closer attention. Both Penney and Lupton (1961) and Spence and Segner (1967) used candy as their reward and noise as their punishment. It is possible that receiving a candy reward served as a more distracting stimulus in both these studies than did the noise punishment. If this were indeed the case, one would expect the R condition to lead to slower acquisition than the R-P condition, and the R-P condition in turn to lead to slower acquisition than the P condition. These were the results of the Penney and Lupton (1961) study, and while the R-P and P condition differences were not significant in the Spence and Segner (1967) study, the trend of the data was in the direction of slower acquisition in the R-P condition.

A minor point of disagreement with a previous study (Meyer and Offenbach, 1962) centers around the complexity variable. The Meyer and Offenbach (1962) findings indicated no significant differences among reinforcement conditions at the level of task complexity that included only one irrelevant dimension. The present study found R-P condition superiority at both the one irrelevant dimension level and the three irrelevant dimension level. Meyer and Offenbach would predict, however, that the magnitude of differences between reinforcement conditions would become greater as complexity increases. This trend was evident in the

data although the treatment by complexity interaction was not significant. It is possible that the failure of Meyer and Offenbach (1962) to obtain significant differences at their most simple level of complexity was due to the relatively weak reinforcing strength of their verbal reinforcers. The response contingencies were fully explained to the *Ss* in the present study, whereas they were not in the Meyer and Offenbach procedure. This procedural difference makes it difficult to compare directly the results of the two studies, and could also easily account for the minor disagreement in findings concerning the complexity variable.

The significant main effect attributable to age in this experiment was not expected on the basis of a previous study (Meyer and Seidman, 1960) in which the concept "larger than" was acquired by *Ss* ages four to five and eight to nine at approximately the same rate. It is possible that the age main effect may have been the result of an interaction between the instructions given to the *Ss* and chronological age. The instructions had the intended effect of making more salient cues of the stimuli associated with the various response contingencies. If the instructions were either understood or retained less well by the kindergarten children than by the third-graders, the experimental task would have been, in effect, more difficult for the younger children, thus accounting for the main effect of age. This analysis is supported by an examination of Table 1 which shows that the age main effect occurred for the R and P conditions but not for the R-P condition, which depended less on instructions for stimulus saliency than did the other two conditions.

The R condition of the present experiment is an example of operant strengthening of responding in the presence of certain stimuli, and extinction of responding in the presence of other stimuli. In the R-P condition, there is operant strengthening in the presence of certain stimuli, but in the place of extinction of responding in the presence of other stimuli, punishment occurs. The difference between condition R and condition R-P is the difference between extinction and punishment. Holz, Azrin, and Ayllon (1963) have convincingly demonstrated that punishment can be a more effective means of eliminating behavior than can extinction. The present results are considered a replication of that finding.

Following an analysis proposed by Azrin and Holz (1966), it is suggested that the increased responding in the presence of the correct stimulus in situation P does not necessarily indicate that learning of the desired discrimination has occurred: ". . . the elimination of a response by punishment is not known to result in an increase of unpunished responses unless those responses are concurrently under the control of some reinforcement procedure" (Azrin and Holz, 1966, p. 431). It is suggested that the increase in nonreinforced responding to the correct stim-

ulus dimension that occurs in condition P is merely a function of the fact that the *S* is forced to choose on each trial. In the P condition, a decrease in punished responding necessarily results in an increase in "correct" responding. The implication of this analysis of the P condition is that its seeming effectiveness is a function of an artificially contrived experimental situation that is not replicable outside the laboratory. An empirically demonstratable implication of this analysis is that given the same rate of acquisition of a forced choice discrimination in both P and R conditions, the P condition *S*s should be inferior on tests of discrimination reversal or tests of transfer. If punishment is to be used effectively in discrimination learning, the "correct" stimulus dimension must be concurrently under the control of some reinforcement procedure. Punishment is not a method of teaching new behavior; it is a method of eliminating behavior. When new behavior seems to emerge as a result of the use of punishment without concommitant reinforcement, as in the P condition of the present experiment, it is well to examine the situation closely to determine possible sources of positive reinforcement. It is hypothesized that both achieving a "correct" solution, as defined by the *E* in his instructions to the *S*s and complying with the *E*'s demand to make a choice on each trial were possible sources of positive reinforcement which functioned concommitantly with punishment to strengthen and maintain responding to the "correct" stimulus dimension on each trial.

REFERENCES

Azrin, N. H., and Holz, W. C. Punishment. In W. K. Honig (Ed.), *Operant Behavior: Areas of Research and Application*. New York: Appleton, 1966. Pp. 380–447.

Brackbill, Yvonne, and O'Hara, J. The relative effectiveness of reward and punishment for discrimination learning in children. *Journal of comparative and physiological Psychology*, 1958, **51**, 747–751.

Curry, Carolyn. The effects of verbal reinforcement combinations on learning in children. *Journal of experimental Psychology*, 1960, **59**, 434.

Gellermann, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *Journal of genetic Psychology*, 1933, **42**, 207–208.

Hoge, M. A., and Stocking, R. J. A note on the relative value of punishment and reward as motives. *Journal of animal Behavior*, 1912, **2**, 43–50.

Holz, W. C., Azrin, N. H., and Ayllon, T. A comparison of several procedures for eliminating behavior. *Journal of the experimental Analysis of Behavior*, 1963, **6**, 399–406.

Meyer, W. J., and Offenbach, S. I. Effectiveness of reward and punishment as a function of task complexity. *Journal of comparative and physiological Psychology*, 1962, **55**, 532–534.

Meyer, W. J., and Seidman, S. B. Age differences in the effectiveness of different reinforcement conditions on the acquisition and extinction of a simple concept learning problem. *Child Development*, 1960, **31**, 419–429.

Meyer, W. J., and Seidman, S. B. Relative effectiveness of different reinforcement

combinations on concept learning of children at two developmental levels. *Child Development*, 1961, **32**, 117–127.

PENNEY, R. K., AND LUPTON, A. A. Children's discrimination learning as a function of reward and punishment. *Journal of comparative and physiological Psychology*, 1961, **54**, 449–451.

SPENCE, JANET T. Verbal discrimination performance as a function of instructions and verbal reinforcement combination in normal and retarded children. *Child Development*, 1966, **37**, 269–281.

SPENCE, JANET T., AND SEGNER, LESLIE L. Verbal versus nonverbal reinforcement combinations in the discrimination learning of middle- and lower-class children. *Child Development*, 1967, **38**, 29–38.

STEVENSON, H. W., WEIR, M. W., AND ZIGLER, E. F. Discrimination learning in children as a function of motive-incentive conditions. *Psychological Report*, 1959, **5**, 95–98.

TERRELL, G., AND KENNEDY, W. A. Discrimination learning and transposition in children as a function of the nature of the reward. *Journal of experimental Psychology*, 1957, **53**, 257–260.

WARDEN, C. J., AND AYLESWORTH, M. The relative value of reward and punishment in the formation of a visual discrimination habit in the white rat. *Journal of comparative Psychology*, 1927, **7**, 117–127.

WISCHNER, G. J., FOWLER, H., AND KUSHNICK, S. A. Effect of strength of punishment for "correct" or "incorrect" responses on visual discrimination performance. *Journal of experimental Psychology*, 1963, **65**, 131–138.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 326–338 (1969)

The Effects of Age, Stimulus Intensity, and Training Trials on Mediated Stimulus Generalization[1,2]

WILLIAM J. MEYER
Syracuse University

AND

ARTHUR ORGEL
School of Medicine, University of Rochester

Three experiments involving the mediated stimulus generalization (MSG) of verbs scaled for hostile intensity are described. Experiment I contrasted gradients for two age groups (9–2 and 11–8). The younger *Ss* had a higher, but not broader, gradient. Experiment II was derived from the Stimulus Intensity Dynamism (V). One-half of the *Ss* (Group H) were trained to make a lever pull to intensely hostile verbs and generalized down the intensity scale. The remaining *Ss* (Group M) were trained to mildly hostile verbs and generalized up the intensity scale. The Group M gradient was higher than that of Group H and consistent with (V). Experiment III contrasted the MSG gradients for groups receiving 4 and 8 reinforced training trials. As predicted the Group 8 gradient was higher and broader.

Stimulus generalization (SG) is said to occur when an organism which has learned a response to a given stimulus (CS) makes a similar response to other, novel stimuli (S_1, S_2 etc.). The typical procedure involves training an instrumental response to a stimulus (conditioned stimulus) and then presenting test stimuli related to the conditioned stimulus (CS). In order to offset extinction effects during generalization trials the CS is included during the test trials and are called "booster trials."

This paper describes three experiments in which children were conditioned to make an instrumental response (lever-pull) to extremely hos-

[1] This study was supported in whole by the U. S. Public Health Service, National Institute of Child Health and Human Development, Grant No. HD-00879. A part of this paper was presented at the Society for Research and Child Development meeting, 1966. The authors express their appreciation to Mr. Edward Kruse, superintendent of the North Hills Joint Schools, Ross Township, Penna. for his cooperation in providing subjects for this study.

[2] The data were collected while the authors were members of the staff of the Psychology Department, University of Pittsburgh.

tile verbs (kill, murder, etc.). Generalization trials consisted of the random presentation of three clusters of five verbs each, whose average intensity-values were progressively less than the intensity-value of the training verbs. Booster trials were interspersed among the test trials. The procedure is consistent with those used in studying primary stimulus generalization and should, therefore, be sensitive to the same variables. The experiments are based on the rationale that the stimulus words are associated with systems of mediating reactions. The mediational systems in turn provide stimuli which can be associated with a variety of instrumental responses, such as a lever-pull. Different stimulus words that occupy approximately the same intensity-position are assumed to be associated with essentially the same cognitive or affective mediating reaction: Hence they can be used to establish an association between a given instrumental response and the mediating processes involved. Systematic changes in the degree of association between stimulus words and the mediating reaction should generate a generalization gradient of the instrumental response.

In an effort to demonstrate that MSG, as defined here, is influenced by the same independent- and subject-variables that influence PSG, the research program focused on age and sex differences, stimulus intensity, number of reinforcements during training, and autonomic reactivity. This paper includes the experiments on age differences, stimulus intensity, and number of training trials.

EXPERIMENT I

This experiment examined the effects of age differences on mediated gradients of generalization. Studies by Mednick and Lehtinen (1957), and Tempone (1965, 1966), using physical stimuli (a spatial array of lights) report steeper generalization gradients among their older *Ss*. Mednick and Lehtinen interpret their results as showing that younger children respond more intensely to stimuli whereas Tempone contends that younger children discriminate less adequately. Despite these differences in interpretation (they will be examined more fully later in this paper), it was expected that the younger *Ss* would respond more intensely to the generalization stimuli.

Method

Apparatus. The apparatus for the MSG research program was designed to minimize the role of the experimenter (*E*). The instrument system was fully automated, with the exception that *E* gave oral instructions to *S* before retiring to another part of the laboratory where he was out of view. Timers, relays, and other equipment were shielded from *S*'s view by

a 3- × 5-foot plywood panel which housed a 1- × 2-foot milk glass screen upon which the stimuli were projected. The screen was located 30 inches from *S*. Immediately in front of the *S* was a semicircular box, housing a 4-inch long lever which could be pulled a distance of 12 inches. The lever was returned to its original position by means of a two-pound spring. Reinforcers were dispensed automatically by means of an M & M candy dispenser (Davis Scientific Instruments, Model MMD-1). Amplitude of response, measured in millimeters, was recorded by means of a Sanborn Recorder (Model No. 954A-100). The stimuli were projected by means of a Kodak 100 Projector located 4½ feet behind the milk-glass screen. Trial duration, intertrial interval, and dispensement of reinforcers was controlled by a Gerbrands Program Timer (Model No. 1001). All equipment was housed in an 8 × 18-foot mobile laboratory.

Stimulus words. An initial list of 216 verbs, judged to connote hostility, were selected from the Thorndike-Lorge word list. All the selected verbs had a frequency of between 20 and 100 occurrences per million words. Each verb was typed in capital letters at the top of an 8½ × 3-inch piece of white, unlined paper. Immediately below the verb was a five-point scale with the following labels identifying each point: Most Bad—Very Bad—Somewhat Bad—Least Bad—Not Bad. The 216 verbs were randomized and combined into booklets of 100 and 116 words each and were presented in counterbalanced order on two successive days. *Ss* were 96 fourth-grade children (54 males and 42 females) and 115 sixth-grade children (60 males and 55 females). The mean IQ of the total sample was 114.6.

Median scale-values and semi-interquartile ranges Q_1–$Q_3/2$ for each of the 216 verbs were determined.[3] Test-retest analysis indicated a reliability of .90. There were no statistically significant sex or grade differences, nor was the Sex × Grade interaction significant ($F = 2.5$; $df = 1$ and 207; $p > .05$; $F = > 1$; $df = 1$ and 207; $p > .05$; $F = 1.7$; $df = 1$ and 207; $p > .05$). Words not understood and/or read without help by one percent of the sample were deleted. From among the remaining words, one group of eight words was selected for training-stimuli and three groups of five words were used as test-stimuli. Groupings were based on similarity of median scale values and the magnitude of the semi-interquartile range. Table 1 summarizes the pertinent statistical information on the stimulus words. The procedure yielded an ordinal scale. It is assumed that the average values of the four intensity levels are discriminable and psychologically meaningful.

To determine the validity of the resulting scale, two experiments were

[3] Semi-interquartile ranges were computed because of the extreme skewing of the distributions for each word.

TABLE 1

MEDIAN SCALE-VALUES AND SEMI-INTERQUARTILE RANGES OF STIMULUS WORDS

Word	Median	Semi-interquartile range	
Level 4			
Shot	4.82	.31	
Killed	4.86	.29	
Hanged	4.81	.37	
Poisoned	4.76	.61	$M = 4.70$
Drowned	4.60	.90	
Murdered	4.64	.72	
Knifed	4.55	.86	
Gunned	4.58	.72	
Level 3			
Flooded	3.47	.89	
Injured	3.50	.71	
Invaded	3.50	.81	$M = 3.48$
Fired	3.50	.73	
Arrested	3.44	.79	
Level 2			
Frightened	2.63	.80	
Defeated	2.63	.79	
Troubled	2.63	.73	$M = 2.63$
Soiled	2.62	.75	
Snapped	2.65	.88	
Level 1			
Left	1.90	.72	
Stared	1.83	.83	
Fooled	1.85	.69	$M = 1.62$
Yelled	1.79	.63	
Frowned	.74	.42	

conducted (details are available in mimeo form, Meyer and Orgel, 1965). In the first experiment, sixth-grade children were presented with a series of 3-inch $\times$ 5-inch cards, on which were the four pronouns "I," "We," "You," and "They" and either a "mildly hostile" (Median scale value = 2.21, $SD = .35$) or an "intensely hostile" (Median scale value = 3.98, $SD = .47$) verb. Within each intensity group, one-half of the *Ss* were reinforced for using "I" and the remaining *Ss* were reinforced for using "They." Consistent with the scale values, the "I"–intensely hostile group conditioned least rapidly. The base rate data showed no difference in frequency of "I" versus "They" responses for the mildly hostile verbs but a substantial difference for the intensely hostile verbs.

The second study employed the apparatus developed for the MSG experiments (Meyer and Orgel, 1965). Essentially, this study involved a

random presentation of the stimulus words shown in Table 1. Twenty sixth-grade children were instructed to pull the handle a distance commensurate with the degree of "badness" of the verb. Two practice trials, involving a very hostile verb and a weak verb, served to define the end points. A rank correlation (rho) between the amplitude scores and the paper-and-pencil scale-values for each verb was .88.

Procedure

Children were seen individually in the mobile laboratory. Each of the stimulus words was photographed and mounted for use with the Kodak 100 Projector. When projected on the screen, the words were approximately 1 inch high. Subjects were given the following instructions: "We are interested in how children react to different words. When this red light comes on (point), a word is going to appear here (point to screen), for a short time. For some words you should move this handle (demonstrates); you must decide if the word is one for which you move the handle. If you do move the handle, you must move it while the word is on the screen and then you should move it to the end of the tract (point). Sometimes you will get an M & M candy for moving the handle and sometimes you will not. Keep your hand on the handle, so that you can move it quickly. Pull the handle on the first word ('eat') so that you can see how it works. Any questions?"

The training or conditioned stimuli were eight intensely hostile words to which the subject was required to pull the lever the full distance. If a full response was made, the *S* received an M & M candy. All *S*s received the same eight training trials and the same series of 15 randomly-ordered test-stimuli (five words at each of three intensity levels). The only constaint on randomization was that no more than two words of identical strength appear consecutively. Interspersed among the test stimuli were 16 of the original training stimuli, which served as booster trials. Each of the eight training stimuli appeared twice during the generalization trials with the constraint that no more than two occurred within a block of five presentations. All full-responses to the booster trials were reinforced. Thus, there was a total of thirty-nine trials: Eight training trials, fifteen test trials, and sixteen booster trials. Each word-stimulus was exposed for a period of 2 seconds; the intertrial interval varied randomly (1, 1½, or 2 seconds) to prevent a rhythmic pulling of the lever. The entire session required approximately 20 minutes. All subjects seemed to understand the instructions and to perform the task appropriately. The design of experiment I involves one variable between groups (fourth-grade vs sixth-grade) and one variable within groups (test stimuli).

Subjects. A total of 18 fourth-grade ($M_{ca} = 9\text{–}2$) and 18 sixth-grade

($M_{ca} = 11\text{–}8$) children served as *Ss*. The children were from a suburban middle-class school district. An approximately equal number of males and females were included in the sample.

Results

An examination of the average number of reinforcements received during training shows averages of 7.4 and 7.6 for grades 4 and 6, respectively. This difference is not statistically significant ($t < 1$). Thus, the instructions were equally effective for the groups and both groups entered the generalization trials with equivalent response strength. A comparison of mean response amplitudes during the training trials for each group were also not statistically significant ($t = < 1$).

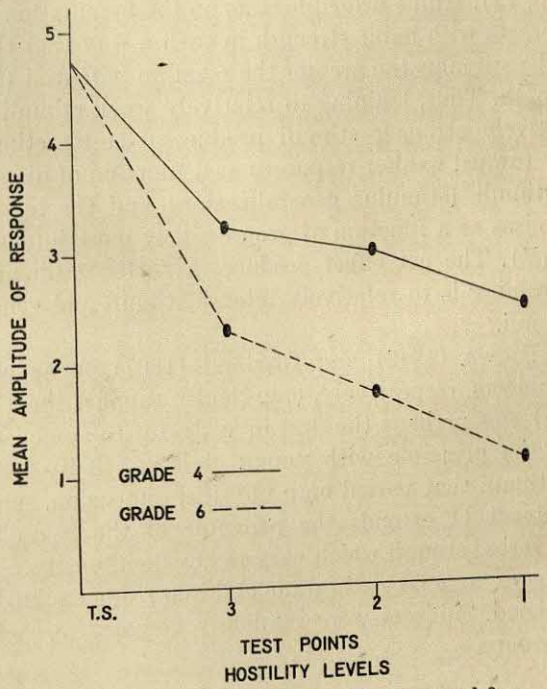


FIG. 1. Gradients of MSG for Grades 4 and 6.

Average response amplitudes were determined for each of the test points. Figure 1 shows the generalization gradients for each group. A repeated measures analysis of variance (Edwards, 1964) involving the three test points resulted in significant differences between age groups in overall level of responding ($F = 8.4$; $df = 1$ and 34; $p < .01$). As anticipated from other studies, the gradient of generalization is higher for the

younger, as opposed to the older, children. An analysis of the gradient effect was not statistically significant ($F = 2.9$; $df = 2$ and 68; $p < .07$). The gradient approaches statistical significance and is the only one of our numerous samples that was not clearly significant. The interaction of stimulus intensity and age was also not statistically significant ($F = < 1$).

EXPERIMENT II

In his discussion of the principle of the Stimulus Intensity Dynamism (V), Hull stated: "The general trend of the joint sEr gradient originating at a weak stimulus intensity and generalizing toward stronger stimulus intensities has a smaller tendency to a downward slope than that extending in the opposite direction between the same stimulus intensities." (Hull, 1949, p. 73). Hull's principle rests on the assumption that stimulus intensity interacts with habit strength in such a way that the greater the intensity of the stimulus the greater the reaction potential (sEr), per unit of habit strength. Thus, training to relatively weak stimuli and generalizing to relatively stronger stimuli produces counteracting forces—or: The tendency toward weaker responses as a function of distance from the conditioned stimuli (stimulus generalization) and the tendency toward stronger responses as a function of progressively more intense generalization stimuli (V). The net effect produces a flatter gradient than would occur when training is to relatively intense stimuli and generalization is to weaker stimuli.

Studies by Brown (1942) and Hovland (1937), using adult humans and rats as subjects, respectively, empirically support the (V) principle. Spiker (1956a) was perhaps the first investigator to demonstrate the applicability of the principle with young children (3–9—5–8). All three studies used stimuli that varied on a physical dimension, such as light or sound. Experiment II extends the principle of the Stimulus Intensity Dynamism to verbal stimuli which vary in hostile intensity.

Subjects. The *S*s were 60 sixth-grade children from a suburban, middle-class neighborhood. Subjects were randomly assigned to each of the two experimental groups.

Procedure

With one exception, the procedures used in Experiment II were identical to those described for Experiment I. Group H was trained to the high intensity words and generalized down the scale; this group, in effect, replicates those in Experiment I. Group M was trained to mildly hostile stimuli and tested for generalization up the scale. The training stimuli for Group M involve the five words at level 1 (Table 1) plus the words

bothered, splashed, and barked (the median scale values and semi-interquartile ranges for these words are, in order: 1.71, .71; 1.55, .62; and 2.00, .84). Three words were deleted from the level 4 group (hanged, drowned, murdered) so that only five words were available at that intensity level. Both groups received the same instructions.

Results

Average response-amplitudes for each test-point and for the conditioning trials were determined. One corollary of the Stimulus Intensity Dynamism (V) is that performance during the conditioning trials should be superior for Group H. This corollary rests on the assumption that the greater the intensity of the stimuli, the stronger the instrumental response. As shown in Figure 1 this prediction was not supported, in fact the mean response amplitudes were reversed.

Figure 2, which has double entries on the abcissa, shows the gradients for each group. Inspection of the curves suggest that in general Group M made stronger responses at each test point and that the slope of its grad-

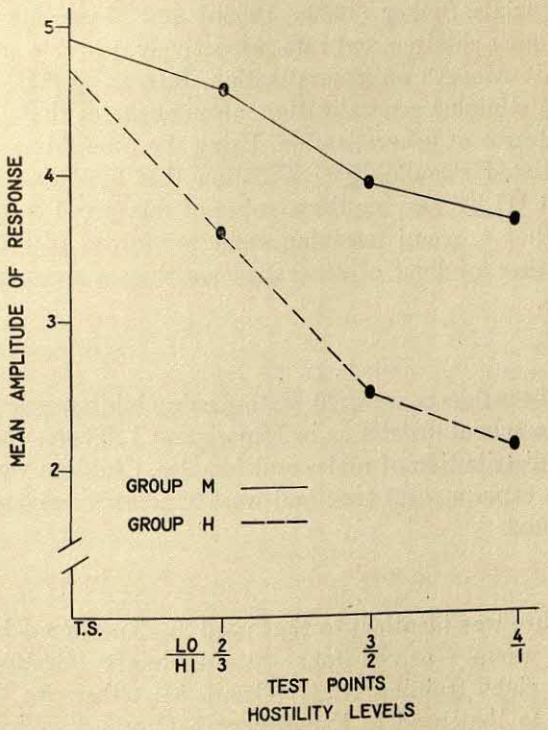


FIG. 2. Gradients of MSG for high vs low intensity training.

ient is less steep. A repeated measures analysis of variance, following the procedures described by Edwards (1964), shows the gradients differ significantly in height ($F = 15.8$; $df = 1$ and 58; $p < .01$). This finding means that, on the average, Group M pulled the lever harder at each test point. The interaction between intensity groups and test stimuli is not statistically significant ($F = 1.2$; $df = 2$ and 116; $p > .05$). The effect attributable to stimuli is also statistically significant ($F = 27.2$; $df = 2$ and 116; $p < .01$), indicating that the combined gradient deviates significantly from the horizontal.

EXPERIMENT III

A second variable assumed to influence generalization gradients is the strength of the learned response (Spence, 1937). Accordingly, if motivation level remains constant but habit strength is permitted to vary, groups with the stronger habit strength should generalize more; that is, their gradient should be higher. Unlike V, which leads to the prediction of differential slopes of generalization gradients, formulations about initial habit strength suggest only a higher gradient for groups receiving more training trials. Spiker (1956a, 1956b) and Margolius (1955), using as subjects, young children and rats respectively, provide support for the effects of habit strength on generalization. Razran (1949), in his review of research on stimulus generalization, also concludes that habit strength is related to degree of generalization. Using the procedures developed for studying mediated stimulus generalization this hypothesis is examined in Experiment III by varying the number of reinforced training trials. It is predicted that a group receiving eight reinforced training trials will evidence a higher gradient of generalization than a group receiving four training trials.

Method

Subjects. The subjects were 40 sixth-grade children selected from the same suburban school district as in Experiment I. There was an approximately equal distribution of males and females. Children were naive with respect to the experimental task and were randomly assigned to the two treatment groups.

Procedure

The procedure was identical to that used in Experiment I with one exception: One group received four training trials (Group 4) and the second group eight training trials (Group 8). Otherwise the procedure was identical to that used in Experiment I. Group 4 subjects were conditioned to the following words; shot, killed, hanged, poisoned. Following

training, both groups were exposed to the same test stimuli and booster trials (including the words deleted for Group 4: drowned, murdered, knifed, gunned).

Results

Despite the difference in number of training trials, response strength at the conclusion of the conditioning trials for each group is almost identical ($t > 1$). Examination of Fig. 3 suggests that the gradient for Group 8 is both higher and broader than for Group 4. A repeated measures analysis of variance indicates that overall response strength is signifi-

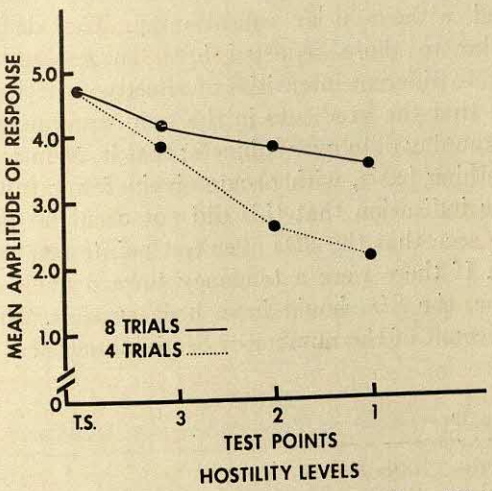


FIG. 3. Gradients of MSG following 4 training trials and 8 training trials.

cantly greater for Group 8 ($F = 17.7$; $df = 1$ and 38; $p < .01$) and supports the hypothesis with respect to the heights of the gradients. The Treatments $\times$ Intensity interaction is statistically significant ($F = 9.0$; $df = 2$ and 76; $p < .01$), indicating that the Group 8 slope is broader (less steep) in contrast with Group 4. The combined curves are also statistically significant ($F = 28.4$; $df = 2$ and 76; $p < .01$).

DISCUSSION

The results of the three experiments permit the tentative conclusion that MSG is influenced by the same variables as PSG. The gradients generated in each of the experiments are consistent with the assumption that the strength of an instrumental response (lever-pull) is related to stimulus intensities; in this case, the affective (hostile) intensity-values of verbs. Verbs within the training series and within each of the three test-

levels varied in their position on an affective intensity dimension and, therefore, in the strength of their mediating reactions. As the mediating reactions decreased in strength, the magnitude of the instrumental response showed a corresponding decrease. We cannot assure, however, that the gradients resulted strictly from affective reactivity, since *Ss* first had to understand the connotative meanings of the words. Indeed, the training trials probably alerted the *Ss* to the nature of the stimuli and the dimension necessary for responding on the test-trials. The *most likely* dimension, however, was "badness," since the explicit meaning of the verbs varied considerably. Support for this position derives from Van Wormer (1966), who recorded galvanic skin responses to the stimulus-verbs, presented without prior conditioning. The GSR gradients are strikingly similar to those reported here, suggesting that the word-groupings generate different intensities of affective reaction.

It is possible that the gradients in the two experiments resulted from an artifact of grouping data over subjects; that is, *Ss* may have responded on an all-or-nothing basis, with progressively fewer full responses over test points. One indication that this did not occur is shown in Table 2 where it can be seen that the *SD*s over test points and samples are relatively constant. If there were a tendency toward not responding at intensity level one, the *SD* should have been smaller. This conclusion is supported by a count of the number of *Ss* giving either a full response or

TABLE 2
STANDARD DEVIATIONS AT TEST POINTS FOR TREATMENT GROUPS

Stimuli	Fourth-grade	Sixth-grade	Group 8	Group 4	Group M	Group H
T.S.	.3	.6	.4	.7	.2	.4
1	1.4	1.8	1.2	1.3	1.1	1.2
2	1.6	1.5	1.6	1.4	1.2	1.7
3	1.8	1.4	1.5	1.6	1.5	1.9

no response to the test-stimuli. The results indicated a maximum of 20% falling into either category, with the majority of these full responses. Thus it seems reasonable to conclude that the gradients are not artifactual but reflect response-strengths associated with the stimuli.

The significant age-effect found in Experiment I is partially consistent with results reported by Mednick and Lehtinen (1957) and Tempone (1965, 1966). Thus, our younger *Ss* pulled the lever significantly harder overall, but the gradients for the two groups were essentially parallel. That the younger *Ss* provide a higher gradient could mean, as Tempone (1966) argues, that they less adequately discriminated among the test

points than the older *Ss*. This position reflects the "stimulus-generalization-as-the-inverse-of-discrimination" position discussed by Prokasy and Hall (1963). Recall, however, that on the paper-and-pencil ratings, there were no age-differences; that is, the younger children did not rate the verbs higher. It may be that the higher gradients reflect a "failure-to-discriminate," but this explanation is not supported by the rating-scale data.

An alternative explanation, consistent with that proposed by Mednick and Lehtinen (1957), is that younger children generally react more strongly to stimuli. Reactions to the rating scales were constrained by the structure of the scales and the instructions, but these constraints were not present with the lever-pull task. Since there were no age differences on the paper-and-pencil ratings of the stimulus-words, it would appear that the higher gradient can be attributed to the greater reactiveness of younger, as opposed to older, children.

Experiment II examined the Stimulus Intensity Dynamism (V) as it relates to gradients of MSG. It was hypothesized that the group trained to low-intensity verbs (Group M) would generate a gradient which was higher and flatter than the group trained to high-intensity verbs (Group H). This prediction was not entirely supported by the data; the Group M gradient was higher but the slope did not differ significantly from that of Group H. It cannot be argued that the essentially parallel slopes seriously question the validity of (V) because no claim can be made for the units of the stimulus scale. It is our contention, on the contrary, that the greater height of the Group M gradient effectively supports (V). Thus among sixth-grade children there is a relationship between response strength and stimulus intensity, where the stimuli are words scaled for hostility.

The final experiment supports the hypothesis that habit strength influences generalization gradients. Consistent with studies using physical stimuli, the greater the habit strength the higher the generalization gradient. It is also apparent from the data, that the slope of the gradient is influenced by the number of reinforcements received during training. Both the slope and height differences are contrary to the Prokasy and Hall (1963) position that generalization is the inverse of discrimination. It can be argued that more reinforced training trials should improve discriminability generating, therefore, a steeper and lower generalization gradient.

The procedures described here would seem to hold promise for a variety of studies of the relationship between stimulus intensity and instrumental responses. An excellent source of possible experiments with children is provided by Buss (1967).

REFERENCES

BROWN, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *Journal of Comparative Psychology*, 1942, **33**, 209–226.

BUSS, A. H. Stimulus generalization and the matching principle. *Psychological Review*, 1967, **74**, 40–50.

EDWARDS, A. *Experimental Design in Psychological Research*, N. Y.: Holt, 1964.

HOVLAND, C. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensity of tone. *Journal of Genetic Psychology*, 1937, **51**, 279–291.

HULL, C. L. Stimulus intensity dynamism (V) and stimulus generalization. *Psychological Review*, 1949, **56**, 67–76.

HULL, C. L. *Principles of behavior*, New York: Appleton, 1943.

MARGOLIUS, G. Stimulus generalization of an instrumental response as a function of the number of reinforcement trials. *Journal of Experimental Psychology*, 1955, **49**, 105–111.

MEDNICK, S., AND LEHTINEN, L. Stimulus generalization as a function of age. *Journal of Experimental Psychology*, 1957, **53**, 100–183.

MEYER, W. J., AND ORGEL, A. *Mediated Stimulus Generalization in Children: Final Progress Report* (NICHD Grant HD-00879), 1965.

PROKASY, W. F., AND HALL, J. F. Primary stimulus generalization. *Psychological Review*, 1963, **70**, 310–322.

RAZRAN, G. Stimulus generalization of conditioned responses. *Psychological Bulletin*, 1949, **46**, 337–365.

SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, 1937, **44**, 430–444.

SPIKER, C. C. The effect of number of reinforcements on the strength of a generalized instrumental response. *Child Development*, 1956a, **27**, 37–43.

SPIKER, C. C. The stimulus generalization gradient as a function of the intensity of stimulus lights. *Child Development*, 1956b, **27**, 85–98.

TEMPONE, V. J. Stimulus generalization as a function of mental age. *Child Development*, 1965, **36**, 229–235.

TEMPONE, V. J. Mediational processes in primary stimulus generalization. *Child Development*, 1966, **37**, 687–696.

VAN WORMER, JANE, *The relationship between mediational activity and the intensity of the galvanic skin response*. Unpublished Master's Thesis, University of Pittsburgh, 1966.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 339–354 (1969)

The Effect of Age, Complexity, and Amount of Contour on Pattern Preferences in Human Infants[1]

BERNARD Z. KARMEL[2]
University of California, Riverside

Two-dimensional grid patterns with black and white elements of four different sizes (⅛-, ½-, 1-, 2-inch checks) and two different arrangements (random or redundant) were presented to 13-, and 20-week-old infants using a complete paired comparisons design. Infants preferred (spent more time looking at) patterns with greater amounts of contour thus verifying earlier reports relating contour information to preferences. Preferences decreased as a pattern began to contain great amounts of contour (very small elements). Older *Ss* preferred patterns with greater amounts of contour compared to younger *Ss*. An inverted U-shaped function of preference to the square root of the amount of contour appeared to be the best description of this looking behavior as well as shifts in this looking behavior with age. Describing these preferences and shifts in preference with age in terms of a "complexity" continuum did not prove useful for these data.

Karmel (1966a, 1966b, 1969) used a modified visual cliff with two surfaces equally shallow for rats and chicks and used a Fantz-type looking box and paired comparisons paradigm for human infants to test pattern preferences. He showed that *Ss* descended to (rats and chicks) or spend a greater time fixating (human infants) a pattern with the greater amount of contour[3] whether this contour was arranged in a random manner or a redundant (checkerboard) manner. Further, it was found that such preferences decreased as the amount of contour in the pattern became very great. That is, preferences decreased as the pattern began to contain greater detail, approaching, but not exceeding, the acuity threshold.

[1] The author thanks M. Bergman, M. Cohen, L. Dettweiler, and P. Miller for assistance during various portions of this experiment.

[2] Data from infants for this study were collected while the author was a Visiting Assistant Professor at the Institute of Child Development, University of Minnesota. The use of the facilities there was greatly appreciated. Portions of the data were collected while the author was supported by a National Institutes of Mental Health postdoctoral fellowship MH-25, 145-01 at the University of California, Riverside.

[3] Amount of contour is defined as the sum of the lengths of all black-white transitions contained in a pattern (Berlyne, 1958). This measure can be calculated similar to "total perimeter" as used by Attneave (1957).

Preferences did not depend on judges' ratings of "complexity" value or on "complexity" value determined by information theoretic measures. Studies by McCall and Kagan (1967), and Salapatek and Kessen (1966) indicated or supported similar conclusions for fixation behavior by infants either with respect to the lack of involvement of "complexity," or with respect to the involvement of contours.[4]

Brennan, Ames, and Moore (1966), on the other hand, have convincingly demonstrated that older infants tend to look at patterns with smaller checks. These developmental shifts to smaller elements, and therefore patterns with more elements in them, tend to support interpretations in terms of preferences for increased "complexity" as defined by number of checks in a checkerboard pattern. A similar interpretation of preferences has been made by Munsinger and Kessen (1966) in terms of preference shifts in "cognitive" uncertainty with the uncertainty linked to the "complexity" or information value of a form, both defined by the number of angles of a random polygon.

However, if looking time increased for greater amounts of contour regardless of the arrangement, "complexity," or uncertainty associated with that contour and decreased as an acuity threshold is approached, then a change in the pattern processing abilities resulting in greater acuity could, in turn, shift looking toward patterns containing greater and greater amounts of contour but not necessarily more "complex" or uncertain patterns. Such a shift in pattern processing ability with age in human infants is apparent from the data reported by Fantz, Ordy, and Udelf (1962) and by Poresky and Doris (1967) using the eye fixation technique. The purpose of this paper is to determine whether a shift in preference behavior is best described by a shift to patterns along a "complexity" continuum or along an "amount of contour" continuum.

METHOD

Subjects. Twenty-eight infants divided into two age groups were used. Infants' ages averaged 13 weeks, ranging between 11 and 14 weeks in the younger group, and, 20 weeks, ranging between 18 and 22 weeks in the older age group. Parents of newborn babies listed in the "Birth Announcement" section of the local newspaper were contacted by letter followed by a phone call. Participating parents were paid $3.00 with approximately 35% of the parents contacted volunteering to participate using this pro-

[4] Although Berlyne (1958) originally found that one property of preferred patterns seemed generally to be the greater amount of contour inherent in the pattern, he related preferences to greater "complexity" with "complexity" defined by the amount of contour within the pattern. This definition of "complexity" thus confounded both amount of contour and "complexity."

cedure. Mothers were asked to bring their infants to the University for testing. Only *Ss* completing all portions of the task are reported here. Approximately 15 *Ss* were eliminated due to failure to complete the task, equipment failure during testing, or obvious visual abnormalities such as crossed eyes.

Stimulus materials. Eight two-dimensional grid stimulus patterns (Fig. 1) were constructed using black and white square elements. Two different arrangements, random or redundant, and four different element sizes (⅛-, ½-, 1-, and 2-inch checks) were used to construct eight stimuli in which the total area of pattern was held constant (12 by 12 inches) but the internal arrangement or size of the elements varied. All patterns were constructed on white poster board and were then photographed using high

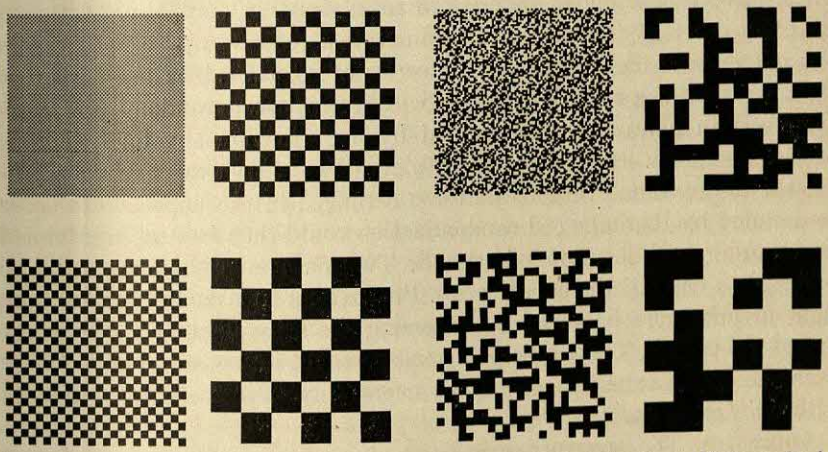


FIG. 1. Stimulus patterns used. Column 1: ⅛-, ½-inch redundant; Column 2: 1-, 2-inch redundant; Column 3: ⅛-, ½-inch random; Column 4: 1-, 2-inch random.

contrast copy film except for the ⅛-inch redundant and ⅛-inch random patterns. The ⅛-inch redundant pattern was produced from commercially available visual aids (Chart-Pak Company) and was then photographed and reduced. The ⅛-inch random pattern was produced by pasting 16 of the ½-inch random patterns together, photographing the conglomerate with high contrast copy film, then reducing the negative to the desired dimension equating it with the other seven stimuli in square area. Use of high contrast copy film and photo negative retouching eliminated the fine lines that appeared at the borders between adjacent like-color areas in the original stimuli. Thus, when projected, all negatives produced white forms on a black background all with an equal square area. The total luminous flux was constant, since the total white was identical in each pattern (50% of a 12-inch by 12-inch area). The brightness from the

patterns averaged 19 foot-lamberts as measured by a Spectra-Brightness Spot Meter, Photo Research Corp., Hollywood. For the larger checked patterns this figure was arrived at by using the average reading for a spot in the black region and one located in the white region of some portion of the pattern.

Presentation order. Since total duration of fixation within a trial is known to decrease by some function over the course of presentations of stimuli to infants, seven different comparison pair series of the 28 possible pairs of stimuli were generated to partially randomize pair position over trials. The series were constrained so that a complete set of the 28 pairs appeared over the first four trials across the seven series, over the next four, and so on through to the last four trials of the seven series. Thus, any general linear decrease in total duration of fixation within a trial was partially balanced over the entire set of seven series of 28 pairs but not necessarily within a single series of 28 pairs. The initial left or right position of a stimulus pattern within a pair was randomized. However, certain pairs were constrained by not allowing any one stimulus pattern to follow itself in the same place. Two slide trays each containing the appropriate stimulus patterns arranged in the appropriate order as dictated by the balanced randomization could then be preprogrammed as to order and location of stimuli. Two *S*s in each age group were assigned to each of the seven series. Preferences to a particular position, right or left, were balanced by reversing the stimulus tray position for one of the two *S*s. These manipulations balanced any position preferences over the entire experiment but not necessarily over a single individual within the experiment.

Apparatus. The apparatus consisted of a commercially secured infant seat mounted on a table 30 inches by 50 inches. The table was centered in a 9 × 12-foot room of a mobile trailer especially designed as a mobile laboratory (Bergman, 1964). The room was dimly lit by an overhead fluorescent lamp containing a translucent filter to reduce the overall illumination level in the experimental room. Pilot tests showed that both too little and too great illumination levels detracted from the task or upset the baby, respectively. The illumination level with no stimuli present was less than 1 foot-candle. Attached to the table was a dull gray frame on a stand which held a rear projection milk screen 16 inches high by 32 inches wide. The angle of the screen was adjusted from the vertical to offset the angle produced by the inclined infant seat. This was done to enable projection of the whole stimulus form in a plane parallel to *S*'s body as *S* sat in the seat. Two Kodak Carousel slide projectors were used to display the stimuli onto the screen. The height, angle, and distance of the projectors with respect to the screen was adjusted so that two

12 × 12-inch images could be projected. These two fields were 2 inches from the top and bottom of the screen and 1½ inches from each side with a 5-inch separation between stimuli. No partition existed on the screen other than that produced by the distance between stimulus patterns. When *S* was seated in the infant seat his eyes were 17 inches from the nearest part and 23 inches from the farthest part of each stimulus pattern at eye level. The visual angle of a ⅛-inch check varied from 26 minutes (near part of screen) to 19 minutes (far part of screen) of visual arc. Below the milk glass screen was a green cloth which was draped from the screen to the top of the supporting table. This cloth served to hide a closed-circuit TV camera. The cloth also helped to direct *S*'s vision to the projected stimuli by reducing extraneous visual stimuli from the straight ahead. The TV camera was focused on *S*'s face through a 2-inch hole cut in the cloth just below the center of the milk glass screen and directly in line with *S*'s face. The camera was connected to a TV monitor housed in a separate room just off of the experimental room. An observer, naive as to the location of particular stimuli to prevent observer bias, monitored the direction of *S*'s look from the TV screen. The observer's task was to judge when *S* looked at a pattern and whether the direction of look was to the left or to the right stimulus. The main clues to these judgments were facilitated by the wrinkling of the eyebrows and forehead produced when *S* looked upwards at the inclined screen from the normal eye position parallel to the floor. Separated counters, activated by the observer when *S* was judged to be looking at a stimulus, accumulated the total time spent looking at each pattern for each trial to the nearest 0.1 second. If *S* was not attending to either stimulus (i.e., exploring hands, judged to be looking away from the stimuli, eyes closed) during a trial, no looking time was accumulated. The interobserver reliability for total amount of looking to each stimulus of a pair was reported to be .89 (Pancrotz, 1968) or .93 for two observers rating looking behavior of similar aged infants using similar techniques (Cohen, 1968).

Testing procedure. Subjects were brought into the trailer room, usually by *S*'s mother, directly from the out of doors and the task was explained to the parent. The parent was then asked to place her infant into the seat and was then told to strap him in around the waist. This procedure prevented *S* from slipping in the seat and also allowed the distance from *S*'s head to the screen to remain relatively constant throughout the testing period. The door to the trailer remained open while the task was being explained to provide light. Testing was begun after this door to the trailer was closed and all persons were in the observation room. Approximately ½ minute to 1 minute elapsed between the presentation of the first comparison pair and placement of *S* in the infant seat. This time was kept

to a minimum to decrease the total time *S* had to spend alone and unoccupied in the chair. Long delays at this point tended to upset the *S*. The use of the Kodak slide projectors enabled electronic programming control over the trial and intertrial intervals. These intervals were 14 and 4 seconds, respectively. Nothing was displayed on the milk glass screen during the intertrial interval. Testing was continuous for the 28 trials of a complete paired comparison set of the eight stimuli. All attempts were made to complete the prescribed series of comparison pairs. Testing was interrupted when no looking behavior to the stimuli could be ascertained for two consecutive trials due to crying, sleep, or attention to other parts of *S*'s environment such as hands. If an interruption was necessary, the stimulus trays and programming equipment were set to redisplay the previous comparison pair and to continue with the experiment. No more than two halts were tolerated. The *E* terminated the task if a third pause was necessary.

"Complexity" Ordering

A determination of the "complexity" ordering of the stimuli was made to verify the independence or orthogonality of the two stimulus orderings: "amount of contour," and "complexity." Although this manipulation had previously been done (Karmel, 1966b, 1969), the particular patterns used in this study constituted a different sample of random patterns and also included a ⅛-inch random pattern not previously used. Therefore, an independent determination of the "complexity" ordering of the stimuli was made.

Method. Undergraduate students of the University of California, Riverside ($N = 28$) were used. Selection of *S*s was made by randomly telephoning undergraduate students listed in the University directory and requesting their volunteering for " . . . an experiment in vision taking approximately 15 minutes for which participants are given $1." Approximately half of the *S*s contacted volunteered. Presentation of the stimuli followed similar orders, controls, and procedures as outlined for infants. Four *S*s, instead of two were assigned to each series. The *S*s were tested individually and were seated 40 inches from a large 70-inch square screen. The projected size of each pattern was 1 square foot. Thus, the visual angle subtended was approximately ½ that for the infants. Each *S* was asked to select the pattern, " . . . you feel is more complex." Each pair of stimuli was projected as long as *S* wished by allowing *S* to decide when to change to the next pair of patterns. At the end of the last comparison pair the ⅛-inch checkerboard stimulus was projected and *S* was asked, "What is this pattern composed of, what is it made up of?" This last procedure was performed to verify the degree of focus possessed by *S*.

All *Ss* after the comparison pairs were presented saw the smallest elemented stimulus as composed of "little blocks" or "squares" indicating an adequate degree of sharpness of focus for these *Ss* under these conditions.

Results. The frequency of choice for each pattern was computed and rankings of these orders were analyzed for any significant monotonic and nonmonotonic trends using procedures outlined by Ferguson (1965) for nonparametric trend analysis. Completely orthogonal stimulus dimensions to one another are listed in Table 1. Frequency of choice data thus

TABLE 1
"COMPLEXITY" AND "AMOUNT OF CONTOUR" ORDERINGS

	Complexity: Randomness supersedes size of element	Average rank judges[a]	Amount of contour	Inches of contour in pattern
	Order		Order	
least				
↓	1. 2-inch Redundant	1.38	1. 2-inch Random	99
	2. 1-inch Redundant	2.14	2. 2-inch Redundant	144
	3. ½-inch Redundant	3.38	3. 1-inch Random	166
	4. ⅛-inch Redundant	4.82	4. 1-inch Redundant	288
	5. 2-inch Random	3.96	5. ½-inch Random	299
	6. 1-inch Random	5.68	6. ½-inch Redundant	576
	7. ½-inch Random	6.80	7. ⅛-inch Random	1163
	8. ⅛-inch Random	7.84	8. ⅛-inch Redundant	2304
most				

[a] Rank of 8 represented highest possible "complexity."

collected would only reverse the relative position of the ⅛-inch redundant pattern and the 2-inch random pattern (see column headed, "Average rank, judges" in Table 1). Only a monotonic trend was significant for judged "complexity" ($z = 15.69$, $p < .001$) suggesting that a consistent "complexity" ordering was produced with rated "complexity" decreasing first as element arrangement varied (i.e., relative redundancy), and then as element size varied. This finding replicated earlier reports (Attneave, 1957; Karmel, 1966b, 1969) suggesting a hierarchy of information theoretic measures as extrapolated to "complexity" orderings. This apparent two-step process resulted in some conflict when the "most complex" stimulus by one rule opposed the "least complex" stimulus by another rule. The infant looking behavior then were analyzed by both "complexity" orderings using the information theoretic decision rules based on randomness superseding element size as one "complexity" ordering, and judges' ratings of "complexity" as the other.

RESULTS

Verification of Control Procedures

A $2 \times 7 \times 28 \times 2$ factorial analysis of variance of the amount of time spent fixating each stimulus of a pair with two independent factors of Age, and Presentation Series, having 2 *Ss* in each age group assigned to each Presentation Series, and with two repeated measures factors of Trials, and Position (left or right) permitted analysis of the control procedures. For this analysis all interaction terms are included in Table 2

TABLE 2
ANALYSIS OF VARIANCE TABLE FOR CONTROL PROCEDURES

Source	Mean square	*df*	*F*
A (Age)	60.74	1	1.04
B (Series)	68.07	6	1.16
Error (between)	58.67	20	
C (Position)	78.04	1	.83
A × C	13.89	1	.15
B × C	40.73	6	.44
Error (within C)	93.51	20	
D (Trials)	10.33	27	2.58[b]
Linear	138.32	1	30.05[b]
Quadratic	29.98	1	6.11[a]
Cubic	38.98	1	10.28[b]
Residual	71.72	24	.76
A × D	3.73	27	.93
B × D	4.41	162	1.10
Error (within D)	4.01	540	
Linear	4.60	27	1.17
Quadratic	4.91	27	1.25
Cubic	3.79	27	.96
Residual	3.94	459	
C × D	22.97	27	.94
A × C × D	17.03	27	.70
B × C × D	17.60	162	.72
Error (within C × D)	24.38	540	

[a] $p < .05$.
[b] $p < .01$.

except for those interactions containing cells with observations from only two *Ss*. The significance of any such interaction with only two *Ss* per cell would be essentially meaningless (for instance an Age × Presentation Series × Trials interaction). The variance attributed to these interactions were combined with the appropriate error terms.

Decreased attention to patterns over trials. Combined time spent look-

ing at both stimuli on each trial diminished over trials regardless of stimulus comparison pair presented ($F = 2.58$, $df = 27/540$, $p < .01$) as was anticipated by the design.[5] Trend analyses of this general decrease in looking over trials, however, indicated that looking did not decrease uniformly over trials. Although a linear trend is significant ($F = 30.07$, $df = 1/27$, $p < .01$) and accounted for 50% of the variance, the quadratic ($F = 6.11$, $df = 1/27$, $p < .05$) and cubic ($F = 10.28$, $df = 1/27$, $p < .01$) trends significantly accounted for an additional 11% and 14%, respectively.

This decrease in looking time measurements is plotted in Figure 2. Changes in slope of looking at stimuli over trials between trials 1–7 and trials 8–22 suggested a possible dark-adaptation effect interacting with *S*'s attention to the task. This *post hoc* hypothesis is suggested since task

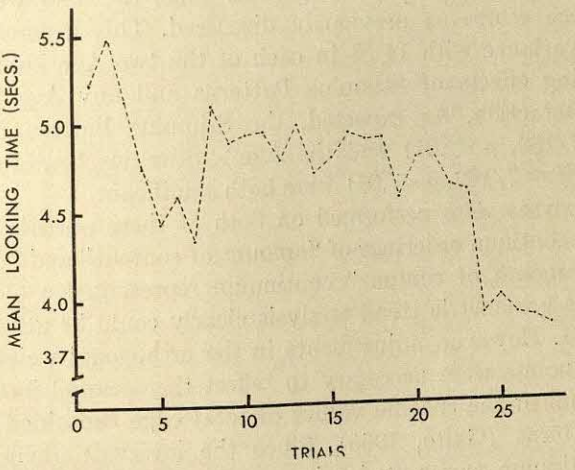


FIG. 2. Decrease in looking time over trials.

onset began relatively soon after room darkening (see procedure), and the rapid decrease followed by a slower decrease seemed to mimic a scotopic-photopic dark-adaptation process. A change in slope from trials 8–22 and trials 23–28 may have resulted from increased crying toward the end of the task. These nonlinear effects further emphasize the multiple interaction of factors, both organismic and environmental, contributing to looking time measurements in infants and necessitated averaging the 7 observations on each stimulus occurring over the 28 trials and using this mean as the dependent measure for effects in the main results.

Position and Age effects. In general, *S*s did not look more to one position (left or right) than to the other position ($F = .83$, $df = 1/20$, n.s.)

[5] $F = 2.64$, $df = 27/378$, $p < .01$, if the variance associated with the Age × Series × Trials interaction is partialed out of the error variance.

or more during a particular Presentation Series than during another ($F = 1.16$, $df = 6/20$, n.s.). Further, no Age differences in time spent looking at stimuli during the course of the experiment indicated that older *Ss* looked for as long a period as younger *Ss* ($F = 1.04$, $df = 1/20$, n.s.). This suggested that any difference between Age groups' responsiveness to particular stimuli could not be due to absolute time spent observing stimuli. The lack of significance of other main and interaction effects (Table 2) suggested that the control procedures employed balanced most extraneous sources of variance over stimulus presentations in the experiment.

Main Results

The lack of a significant Presentation Series effect permitted collapsing scores across Presentation Series groups. Thus the data were analyzed using average scores as previously discussed. This generated a 2×8 analyis of variance with 14 *Ss* in each of the two Age groups and permitted testing effects of Stimulus Patterns and any Age by Stimulus Patterns interaction. As expected, the Stimulus Pattern factor ($F = 8.82$, $df = 7/182$, $p < .01$) and the Age $\times$ Stimulus Pattern interaction ($F = 3.14$, $df = 7/182$, $p < .01$) were both significant.

Trend analyses were performed on both of these partitions using the two distinct continua orderings of "amount of contour" and "complexity." Since the "amount of contour" continuum represented a clear physical continuum, a parametric trend analysis clearly could be utilized (Winer, 1962, p. 353). However, adjustments in the orthogonal weights assigned to stimulus points were necessary to reflect the unequal intervals along the continuum of the specific values of total edge contained in the constructed patterns (Gaito, 1965). Since the intervals along the "complexity" continuum represented only an ordinal relationship, proper trend analyses would be to use a nonparametric trend analysis such as described by Ferguson (1965). This was done. However, in order to allow "complexity" explanations the most powerful statistical test possible, we, for the sake of demonstration, assumed an interval continuum with intervals equal to the mean rank of the judges' "complexity" ratings between stimulus patterns. Thus, the parametric trend analysis for these data along a "complexity" continuum was assumed to be an approximation of a polynomial description of the observed trends, but at the same time, represents some limits for any such polynomial derived to account for the observed variance.

The parametric trend analyses are summarized in Table 3. From a comparison of the proportion of sums of squares accounted for by each component in Table 3, it is obvious that some characteristic of the stim-

TABLE 3

TREND ANALYSES OF TOTAL LOOKING TIME MEASUREMENTS FOR HUMAN INFANTS 13 AND 20 WEEKS OF AGE

		Square root of edge		Actual amount of edge		Judges' "Complexity"	
Source	*df*	*SS*[a]	*F*	*SS*	*F*	*SS*	*F*
Age	1	60.71	1.00	60.71	1.00	60.71	1.00
Error	26	1582.71		1582.71		1582.71	
Stimuli	7	1250.00	8.82[c]	1250.00	8.82[c]	1250.00	8.82[c]
Linear	1	198.43	4.01	65.71	1.40	293.00	10.71[c]
Quadratic	1	1030.43	27.51[c]	1051.86	26.31[c]	105.86	4.66[b]
Cubic	1	.00	.00	96.29	10.51[c]	300.00	18.19[c]
Quartic	1	1.00	.11	14.14	1.31	127.14	5.79[b]
Residual	3	19.86	.54	22.14	.63	424.14	7.98[c]
Age × Stimuli	7	445.00	3.14[c]	445.00	3.14[c]	445.00	3.14[c]
Linear	1	55.86	1.13	99.57	2.13	5.86	.21
Quadratic	1	214.14	5.73[b]	66.43	1.66	2.57	.11
Cubic	1	166.43	19.13[c]	238.86	26.08[c]	2.14	.13
Quartic	1	.00	.00	34.00	3.15	362.14	16.52[c]
Residual	3	8.58	.23	6.14	.26	71.71	1.35
Error	182	3682.00	—	3682.00	—	3682.00	—
Linear	26	1287.00	4.03[c]	1216.57	4.02[c]	711.43	1.54
Quadratic	26	973.71	3.05[c]	1039.43	3.43[c]	591.00	1.28
Cubic	26	226.29	.71	238.29	.79	428.71	.93
Quartic	26	237.86	.75	280.29	.93	571.00	1.24
Residual	78	958.43	—	908.71	—	1381.14	—

[a] Sum of squares is given rather than mean squares to facilitate comparison of proportion of variance attributed to each trend.

[b] $p < .05$.

[c] $p < .01$.

uli along an "amount of contour" continuum seemed to describe the data with fewer parameters than description along a "complexity" continuum. A significant tetratonic trend (Table 4) for the nonparametric analysis of the "complexity" ordering would suggest a similar conclusion. A transformation of the stimulus continuum to the square root of the total inches of contour in the pattern seemed to be the simplest heuristic description of the stimulus factor and its relationships to age and total looking time (see columns in Table 2 under "Square root of edge"). Such a transformation implies an involvement of the area of an element of a pattern in the determination of looking behavior using this paradigm. A "complexity" continuum, either determined by judges' ratings of stimuli or by rules

TABLE 4

Z VALUES OBTAINED USING NONPARAMETRIC TREND ANALYSIS OF TOTAL LOOKING TIME AS FUNCTION OF JUDGES' "COMPLEXITY" RANKINGS OF STIMULI

	Factor		
		Age × Stimuli	
Trend	Stimuli	Younger	Older
Monotonic	2.69[a,c]	2.08[b]	1.69
Bitonic	2.44[b]	0.77	2.70[a,c]
Tritonic	4.63[c]	3.06[c]	3.46[c]
Tetratonic	3.06[c]	0.65	3.65[c]

[a] Using ranks derived from information theoretic measures, i.e., inherent randomness of elements superseding inherent size of elements as a basis for "complexity" orderings, $\alpha = .05$. All other alpha levels remained the same.

[b] $p < .05$.

[c] $p < .01$.

derived from information theoretic measures, could account for the data in either a qualitative or quantitative manner. However, to do so would ignore some simpler relationship between the edge contained in the pattern and subsequent time spent inspecting that pattern.

Slope change of looking behavior with age. Separate analysis for significant trends across stimuli for each age group was performed to determine the nature of the differences found by the Age × Stimulus Pattern interaction. Since "complexity" definitions produced significant residual differences after the 4th order (quartic) trend was tested, further interpretation of the Age × Stimulus Pattern interaction using a "complexity" continuum would be meaningless except as individual comparisons of means between age groups at each stimulus point. Therefore, trends were determined only for the "amount of contour" continuum. The square root transformation of the amount of edge was used for this analysis since this transformation best accounted for the data. This analysis indicated that younger *Ss* showed a significant quadratic trend in looking pattern to the stimuli ($F = 9.10$, $df = 1/13$, $p < .01$), while older *Ss* showed a significant cubic trend ($F = 7.55$, $df = 1/13$, $p < .05$). Figure 3 contains the plot of the Age × Stimulus Pattern interaction. The bold lines at 13 and 20 weeks reflect the quadratic character of younger *Ss*' and the cubic character of the older *Ss*' looking time measurements. The shift in looking behavior along the "amount of contour" continuum with age is apparent from the plotted data. A shift along a "complexity" continuum could only be described in terms of a specific stimulus pattern shift.

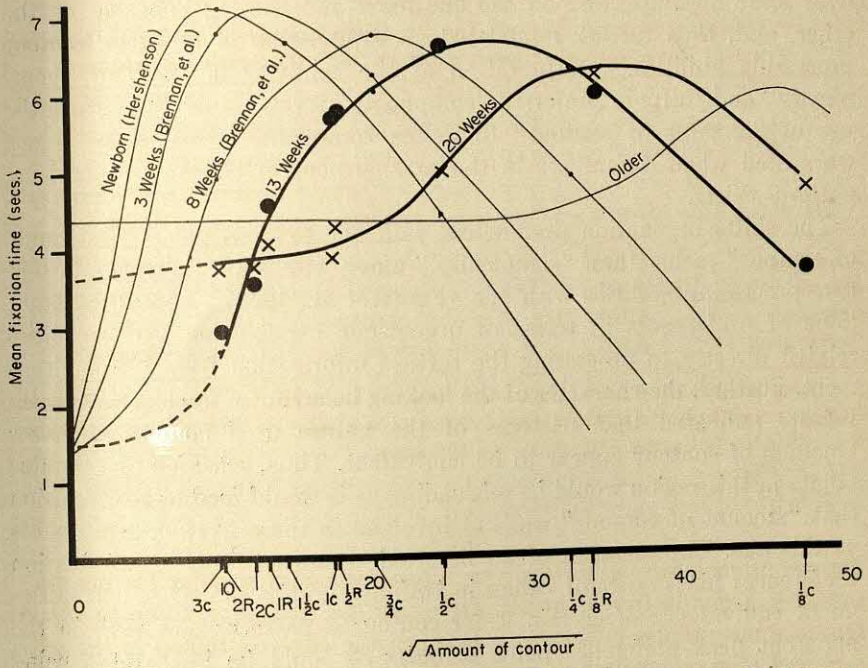


FIG. 3. Mean looking time for each stimulus plotted with respect to the square root of the amount of contour contained in the stimulus for different aged *Ss*. The letters (C) and (R) represent redundant and random patterns, respectively. Number prefix indicates size of element used for construction. (Example: ½C represents the ½-inch redundant pattern.) A lower case (c) represents the redundant character of elements of patterns used by Brennan *et al.* (1966) or Herschenson (1964). The number preceding the lower case (c) would be equivalent to the check size used by these investigators.

DISCUSSION

We conclude that this type of looking behavior decreases with decreasing amounts of contour in the stimulus pattern, and also decreases when a great deal of contour exists in the pattern. A shift in the ability to process edge information as represented in this study by an increase in age is argued to produce a corresponding shift in this apparent inverted U-shaped function of looking behavior with the amount of edge. Since edge increases exponentially with linear change in visual angle, the importance of a square root transformation on the stimulus dimension would imply that the area over which edge occurs would be critical for looking time measurements in tasks using patterned stimuli. Although area of a pattern was held constant in the study reported here, this transformation implied interaction of quantities of contour in the stimulus

over some absolute area on the one hand and looking behavior on the other, and, thus further implicated an explanation in terms of contour processing abilities. Interpretation of the behavior in terms of "complexity" and shifts in preferred "complexity" levels would necessitate the use of "amount of contour" to define complexity. This usage is not warranted when differences in the arrangement of elements comprising patterns exists.

The shifts in fixation preferences with age towards greater "amounts of contour" rather than "complexity," along with known changes in texture processing abilities with age (Fantz *et al.*, 1962), support explanations of preferences in terms of organismic capabilities and capacities related directly to processing the texture information from the environment. Further, the character of the looking behavior in the older, 20-week infants indicated that patterns of the nature used containing lesser amounts of contour appear to be equivalent. Thus, selection of stimulus points in this region would be misleading as it would lead to a conclusion that "amount of contour" was not involved in these fixation preferences at this age. The presence of an inverted U-shaped function of fixation preferences for a range of values in both age groups as well as the subsequent equivalence of certain lesser contoured patterns and shift in the apparent peak-preference point in older *S*s could be interpreted using uncertainty reduction notions of information theory (Munsinger and Kessen, 1966). For such a notion to be meaningful, however, the shift in terms of uncertainty reduction must be related to specific physical visual ambiguities produced by pattern processing differences of this edge information at different ages as opposed to a more cognitive, reduction of uncertainty associated with some set of experiential events or experiences specific to these patterns.

If these conclusions hold, data collected by others should generally fall into this pattern of responding. The thin-lined curves of Figure 3 are hypothetical curves inferred from the looking time data reported by Brennan *et al.* (1966) for 3-, 8-, and 14-week-old *S*s and by Herschenson (1964) for newborns. The distances from *S* to the stimulus patterns used by these investigators were approximately the same as used in the study reported here. However, the size of a pattern and, thus, the total area of pattern was only 6 inches square. To equate for density of edge per area of pattern, the relative position of the stimulus along the abscissa was determined by assuming that the same element size contained in the 6-inch square checkerboard occurred in a pattern 12 inches square. The amount of contour was then calculated on the pattern with the larger area and the square root of this value was used to plot the corresponding

location of these points along the abscissa.[6] The hypothetical curves produced implied by extrapolation of the data to an "amount of contour" continuum and subsequent shifts of looking behavior with age produces the same ordinal relationships of empirically produced preferences to stimulus patterns reported by the investigators.

Differences in the most preferred portions of the "amount of contour" continuum as a function of species (Karmel, 1966a, 1969) would indicate that the basic relationship between "amount of contour" and preference behavior could be maintained with developmental shifts as well as species differences viewed in terms of increased or decreased capacity to process the contour information from the environment. Similarity of shifts in preferences with development as well as similarity of preferences across species tend to detract from explanatory concepts of these behaviors which are species-specific, and tend to shift hypothetical concepts in terms of a common set of parameters, perhaps physiologically based, for this type of behavior. Such an interpretation, based on activity levels of cells in the visual system responsive to edge information has been suggested elsewhere (Karmel, 1969).

Finally, it is suggested that shifts or differences in acuity or other texture processing abilities not only with age (Fantz *et al.*, 1962; Wilson and Riesen, 1966), but also with species-specific optic characteristics (Walls, 1942) or rearing experience (Wilson and Riesen, 1966; Riesen, Ramsey, and Wilson, 1964) may be more important for understanding normal visually dependent functioning than previously surmized.

REFERENCES

Attneave, F. Physical determinants of the judged complexity of shape. *Journal of Experimental Psychology,* 1957, **53,** 221–227.

Bergman, M. A mobile laboratory for research in child psychology. Unpublished M.A. thesis, University of Minnesota, 1964.

Berlyne, D. E. The influence of the albedo and complexity of stimuli on visual fixation in the human infant. *British Journal of Psychology,* 1958, **56,** 315–318.

Brennan, W. M., Ames, E. W., and Moore, R. W. Age differences in infants' attention to patterns of different complexity. *Science,* 1966, **151,** 354–356.

Cohen, L. B. Observing responses, visual preferences, and habituation to visual stimuli in infants. *Journal of Experimental Child Psychology,* 1969, in press.

Fantz, R. L., Ordy, J. M., and Udelf, M. S. Maturation of pattern vision in infants during the first six months. *Journal of Comparative and Physiological Psychology,* 1962, **55,** 907–917.

Ferguson, G. A. *Nonparametric trend analysis.* Montreal, Canada: McGill University Press, 1965.

[6] The points for 14-week infants (Brennan *et al.*, 1966) are assumed to approximate those plotted along the 13-week curve.

GAITO, J. Unequal intervals and unequal n in trend analyses. *Psychological Bulletin,* 1965, **63**, 125–127.

HERSCHENSON, M. Visual discrimination in the human infant. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 270–276.

KARMEL, B. Z. Randomness, complexity, and visual preference behavior in the hooded rat and domestic chick. *Journal of Comparative and Physiological Psychology,* 1966, **61**, 487–489. (a)

KARMEL, B. Z. The effect of complexity, amount of contour, element size, and element arrangement on visual preference behavior in the hooded rat, domestic chick, and human infant. (Doctoral dissertation, The George Washington University, 1966) Ann Arbor: University Microfilms, 1967, No. 66-11, 527. (b)

KARMEL, B. Z. Complexity, amount of contour, and visually dependent preference behavior in hooded rats, domestic chicks, and human infants. *Journal of Comparative and Physiological Psychology,* 1969, in press.

MCCALL, R. B., AND KAGAN, J. Parameters of attention in the infant. *Child Development,* 1967, **38**, 939–952.

MUNSINGER, H., AND KESSEN, W. Stimulus variability and cognitive change. *Psychological Review,* 1966, **73**, 164–178.

PANCRATZ, C. N. Unpublished M.A. thesis, University of Illinois, 1968.

PORESKY, R. H., AND DORIS, J. Brightness and visual acuity in infancy. Paper presented to the Eastern Psychological Association meetings, Boston, April, 1967.

RIESEN, A. H., RAMSEY, R. L., AND WILSON, P. D. Development of visual acuity in Rhesus monkeys deprived of patterned light during early infancy. *Psychonomic Science,* 1964, **1**, 33–34.

SALAPATEK, P., AND KESSEN, W. Visual scanning of triangles by the human newborn. *Journal of Experimental Child Psychology,* 1966, **3**, 155–167.

WALLS, G. L. *The vertebrate eye and its adaptive radiation.* Bloomfield Hills, Michigan: Cranbrook Institute of Science, 1942. (New York, Hafner, 1963, Republished.)

WILSON, P. D., AND RIESEN, A. H. Visual development in Rhesus monkeys neonatally deprived of pattern light. *Journal of Comparative and Physiological Psychology,* 1966, **61**, 87–95.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 355–365 (1969)

Dimensional Shifts and Cue Retention in Retardate Discrimination Learning[1]

THOMAS F. CUNNINGHAM[2]
Oklahoma State University

The theoretical systems of Zeaman, House, and Goodwin and Lawrence were used to predict performance in intradimensional (ID) and extradimensional (ED) shifts and in conditions designed to test the retention of previously acquired cues. Sixty retarded children were divided into six groups and given the same original training on a simultaneous visual discrimination problem. Groups were given either two successive ID or ED shifts after original training. Within each second shift, three types of cue conditions were present: (a) The same condition which had the identical cues and reinforcement contingencies of original training, (b) the New condition in which cues, not experienced previously, were used, and (c) the Reversal condition in which the cues of original training were present but reinforcement contingencies reversed. The findings were: (a) ID-shift performance was superior to ED-shift performance in the second shifts and (b) performance in the Same condition was superior to performance in the New and Reversal conditions. Implications of these findings for the Zeaman and House and Goodwin and Lawrence positions are discussed.

Zeaman and House (1963, p. 166) postulate that a chain of two responses is required to explain retardate discrimination learning: (a) An attentional or observing response to the relevant dimension and (b) an instrumental approach response to the correct cue of that dimension. The probability of observing the relevant dimension (Po) and the probability of approaching the correct cue (Pr) develop as a result of reinforcement and are altered by various transfer operations, e.g., ID and ED shifts. In an ID shift, the relevant dimension of the previous problem remains relevant but new cues of that dimension are introduced. An ED shift involves the substitution of a new relevant dimension for the previously relevant dimension. House and Zeaman (1962) compared the performance of retarded *Ss* in ID and ED shifts in simultaneous two-choice visual discrimination problems and found ID-shift

[1] This research was carried out in partial fulfillment of the degree of Doctor of Philosophy and was partially supported by training Grant HD-44-07 from NIMH. Special gratitude is extended to Jerome Smith, Don Tyrrel, and David Weeks for their aid and constructive criticism. I also thank the children and personnel of Hissom Memorial Center, Sand Springs, Oklahoma.

[2] Now at the State University of New York College at Potsdam.

performance was superior. Bernsberg (1958), and Campione, Hyman, and Zeaman (1965) have reported similar results with retardates. Zeaman and House (1963, p. 193) account for ID-shift superiority by assigning different Po values to the start of ID and ED shifts. The high Po acquired in original training (OT) remains high in an ID shift since the same dimension is relevant in both stages of training. However, transfer from OT is poor and Po is low in an ED shift since a new relevant dimension is used.

House and Zeaman (1963b) investigated retardate learning set formation on 108 two-choice visual discrimination problems. A stimulus pool of four objects provided an opportunity to test for cue retention by manipulating stimulus overlap and reinforcement contingencies on successive problems. No evidence for cue retention was found. Goodwin and Lawrence (1955) reported cue retention in rats on successive ED shifts when the relevant cues of the second shifts were those of OT. Two cue-conditions were used in the second shift problems: (a) The same positive and negative cues of OT were used and (b) the OT positive and negative cues were reversed. Performance in the former condition was superior in the second shift and subsequent ED shifts using a similar paradigm. Goodwin and Lawrence (1955, p. 442) propose a chaining mechanism to account for their findings: (a) An identification of a dimension or set of stimuli and (b) the establishment of preference between stimuli within a set. The acquisition and extinction of the identification response are assumed to occur more rapidly than the establishment and extinction of cue preferences. This assumption provides for cue retention since, in an ED shift, the rapid extinction of the original identification response prevents exposure of the cues of the OT dimension to extinction and permits positive transfer when the OT dimension and cues are made relevant in the second shifts.

The observing response of Zeaman and House and the identification response of Goodwin and Lawrence are similar in function, i.e., each selectively limits the number of cues to which instrumental responses become associated. The Goodwin and Lawrence assumption of a rapidly established and extinguished mediating response is in contrast to the assumption of Zeaman and House (1963, p. 214) that mediating and associative processes may occur at a similar rate. Finally, the emphasis on cue retention in the system of Goodwin and Lawrence has no parallel in the system of Zeaman and House.

The present investigation assessed the relative adequacy of the above mentioned theories to account for both the dimensional shift performance and cue retention of retarded *Ss*. *Ss* were given either two successive ID or ED shifts after OT. Within each second shift, the dimension of OT

was relevant and the cues of that dimension were manipulated to produce three conditions: (a) The *Same* cue condition which had the identical cues and reinforcement contingencies of OT, (b) the *New* cue condition in which cues, not experienced previously, were used, and (c) the *Reversal* cue condition in which the identical cues of OT were present but the reinforcement contingencies associated with these cues were reversed.

According to the Zeaman and House (1963) position, it is hypothesized that ID-shift performance is superior to ED-shift performance in each of the successive shifts. No hypothesis concerning differential performance among the three cue conditions in the second shifts is possible since House and Zeaman (1963a, p. 319) do not postulate retention of cue probabilities (Pr) across problems which involve the learning of new cues. Goodwin and Lawrence (1955) are unable to predict ID and ED-shift differences since their mediating response mechanism operates more rapidly than the rate of aquisition and extinction of overt instrumental associations. However, the Goodwin and Lawrence emphasis on cue retention indicates differences in performance among the cue conditions; i.e., performance in the *Same* cue condition is hypothesized superior to performance in the *New* and *Reversal* cue conditions since the positive cue of OT is the positive cue in the *Same* condition, is absent in the *New* condition, and is used as the negative cue in the *Reversal* condition. Performance in the *New* condition falls between the *Same* and *Reversal* conditions.

METHOD

Subjects. The *Ss* were 88 institutionalized retardates, selected randomly from an MA range of 5½–10½ years, irrespective of past experience and clinical diagnostic category. However, all *Ss* had normal hearing and vision. Twenty-eight of the 88 *Ss* failed to reach acquisition criterion, 24 in OT and 4 in the first ED shift. Sixty *Ss* were used throughout the experiment. These were randomly assigned to each of six experimental groups with the provision that each group had an equal number of *Ss*, and that the range of MAs and average MA of each group were approximately equal. The descriptive statistics of the six groups are presented in Table 1.

Apparatus. The apparatus was a version of the Wisconsin General Test Apparatus (Harlow, 1942), as modified by Zeaman and House (1963). It consisted of a table with a sliding stimulus tray 30 × 12 inches with two circular foodwells 2 inches in diameter centered 12 inches apart. A one-way screen was positioned across the center of the table which permitted observation of *S*'s behavior during testing.

TABLE 1
MA AND CA CHARACTERISTICS OF EXPERIMENTAL GROUPS

	MA (in years)		CA (in years)	
	Mean	SD	Mean	SD
ID-Same	7.2	1.4	15.7	3.3
ID-New	7.3	1.2	14.7	3.4
ID-Reversal	7.2	1.5	13.3	2.7
ED-Same	7.2	1.4	13.2	2.8
ED-New	7.2	1.2	14.8	2.7
ED-Reversal	7.3	1.4	14.5	2.2

The stimuli were painted forms cut from ¼-inch masonite mounted vertically on 4 × 4-inch masonite bases. The height and width of each stimulus was 2 inches. A total stimulus pool of 36 objects was used; six different forms (triangle, circle, square, cross, T, and diamond) in each of six colors (red, blue, yellow, green, black and white).

General procedure. Each *S* was brought into the experimental room, seated before the apparatus, and informed he was going to play the "candy game." Three pretraining trials preceded the initiation of experimental trials. Each pretraining trial consisted of placing candy in full view of *S* in one of the two foodwells and instructing *S* to find the candy. No covering was placed over the foodwells in Trial 1. In Trial 2, two identical plexiglass wedges partially covered the foodwells. The foodwells were completely covered by the wedges on Trial 3. After completion of pretraining, the testing session was begun. A testing session consisted of 25 trials per day. Acquisition criterion was 20 or more correct responses during a daily testing session.

The same procedure was used for all test trials. Before beginning a trial, the stimulus tray was pulled behind the one way screen, and candy reward (M & M) was placed in one of the two foodwells, and the foodwells were covered with the appropriate stimulus objects. A trial consisted of pushing the stimulus tray in front of *S* and permitting him to make a single choice of the stimulus objects. A noncorrection procedure was used. In addition to the candy reward, *E* said "Good" if the response was correct, and "No" if it was not. The position of the correct stimulus was varied according to the Gellerman (1933) series.

Training Stages

Stage 1. *Ss* were divided into six groups of equal size ($N = 10$). These groups were labeled according to the dimensional shifts and cue conditions of the training stages: ID-*Same,* ID-*New,* ID-*Reversal,* ED-

Same, ED-*New*, and ED-*Reversal*. An example of the training sequences for the six experimental groups in terms of the relevant and irrelevant cues used in each stage of training is presented in Table 2.

All *S*s were given OT on a form-relevant problem in which color was varied and irrelevant within trials. Two forms and two colors were chosen independently and randomly from the stimulus pool of six forms

TABLE 2

COLOR (C) AND FORM (F) CUES FOR THREE TRAINING STAGES[a]

Trials	Original Training	
1	+ C_1 F_1 red △	− C_2 F_2 blue ○
2	+ C_2 F_1 blue △	− C_1 F_2 red ○

	ID Shift—no. 1		ED Shift—no. 1	
1	+ C_3 F_3 white T	− C_4 F_4 black □	+ C_5 F_3 green T	− C_6 F_4 yellow □
2	+ C_4 F_3 black T	− C_3 F_4 white □	+ C_5 F_4 green □	− C_6 F_3 yellow T

ID Shift—no. 2 and ED Shift—no. 2

	Same		New		Reversal	
1	+ C_1 F_1 red △	− C_2 F_2 blue ○	+ C_2 F_5 blue cross	− C_1 F_6 red ◇	+ C_1 F_2 red ○	− C_2 F_1 blue △
2	+ C_2 F_1 blue △	− C_1 F_2 red ○	+ C_1 F_5 red cross	− C_2 F_6 blue ◇	+ C_2 F_2 blue ○	− C_1 F_1 red △

[a] A relevant cue was never used in a subsequent problem as an irrelevant cue.

and six colors. Each *S* was randomly assigned one of the form cues as positive with the provision that each of the form cues appeared equally often across *S*s as the positive cue. Each irrelevant color cue was randomly assigned to one of the form cues within a trial, and the two cues appeared equally often with both form cues across trials during training sessions. The same procedure was applied to the relevant and irrelevant cues of each problem for each *S* throughout the experiment. Training consisted of reaching acquisition criterion plus 100 over-learning

trials. The failure criterion for OT was 200 trials without reaching acquisition criterion.

Stage 2. *Ss* in the three ID groups (ID-*Same*, ID-*New*, and ID-*Reversal*) were given the first ID shift in which form was relevant and color was variable and irrelevant within trials. Two new form cues were selected from the four which had not been used in OT. The remaining four color-cues were set in all possible combinations of two and randomly assigned to each *S* as the variable irrelevant cues. For *Ss* in the three ED-groups (ED-*Same*, ED-*New*, and ED-*Reversal*), the first ED shift had color relevant and form variable and irrelevant within trials. The color cues not used in OT were set in all possible combinations of two, and one set randomly assigned each *S*. Two new form-cues were chosen from the form cues not used during OT and served as the irrelevant cues of the first ED shift.

The failure criterion for the first ID- and ED-shift problems was 300 trials without reaching acquisition criterion. One hundred trials of overlearning were given each *S* after reaching acquisition criterion.

Stage 3. The three ID-groups were given a second ID shift and the three ED-groups were given a second ED shift. Form was the relevant dimension and color was variable and irrelevant within trials in both shifts. The ID- and ED-*Same* groups received the same form cues and reinforcement contingencies of OT. *Ss* in the ID- and ED-*New* groups were given two form cues which had not been used previously. *Ss* in the ID- and ED-Reversal groups received the form cues of OT and a reversal of the OT reinforcement contingencies. The irrelevant color-cues were selected for each *S* in the following manner: (a) By eliminating the color cues employed in the preceding problem of Stage 2 and (b) by randomly selecting one of the possible six combinations of the remaining four colors.

RESULTS

The dependent measure used in all analyses was the number of errors to acquisition criterion. A logarithmic $(X + 1)$ transformation of error scores was used to correct for heterogeneity of variance. Mean log errors and standard deviations for each goup in Stage 1 and for the ID and ED shifts of Stage 2 are presented in Table 3.

The performance of the six groups in Stage 1 was analyzed to determine if there were differences among groups prior to the presentation of shift conditions in Stage 2. A one-way classification analysis of variance indicated no effect of groups on performance ($F < 1$). A t test of ID- and ED-shift means in Stage 2 was made by comparing the average performance of the three ID- and three ED-groups. Shift means

TABLE 3
MEAN LOG ERRORS AND VARIANCE FOR STAGE 1 AND STAGE 2

Statistics	Groups					
	ID			ED		
	Same	New	Reversal	Same	New	Reversal
	Stage 1: OT					
Mean log errors	1.14582	0.80429	0.86929	0.84470	0.94997	0.87830
Log variances	.39937	.41470	.24231	.24080	.30762	.22529
	Stage 2: ID Shift no. 1 and ED Shift no. 1					
Mean log errors	.137422			1.07625		
Log variances	.14487			.27493		

and variances for the ID and ED groups were .374, .145, 1.076, and .275, respectively. Since the shift variances were unequal ($F = 1.89$, $df = 29$ and $29, p < .05$), each estimate of variance was used in computing the denominator of the t test. A conservative approximation of the tabulated t value, t' (Steele and Torrie, 1960, p. 81) was used as the criterion measure against which the calculated t value was compared. The t' value corresponds to the tabular t for 29 degrees of freedom. Performance in the ID shift was superior ($t = 5.94$, $df = 29$, $p < .001$).

A completely randomized design with a dimensions (ID, ED) by cues (*Same, New, Reversal*) factorial arrangement was planned to examine shift differences and cue retention in Stage 3. However, the assumption of homogeneity of variance within all six treatment groups was rejected because of the atypical performance of the ID-*Same* group. The ID-*Same* group had zero mean and zero variance; i.e., *S*s responded to the correct cue without error. A Barlett's chi-square test (Steele and Torrie, 1960, p. 347) of homogeneity of variance of the remaining five groups did not lead to rejection of the assumption ($X^2 = 5.91$, $df = 4$, $p > .10$). A single classification analysis of variance was performed on the data of these five groups to obtain an estimate of the common error variance ($s^2_\varepsilon = .24449$, $df = 45$). To make the preplanned comparisons of Stage 3, a series of orthogonal contrasts (Steele and Torrie, p. 215) among treatment means was made using the Least Significant Difference statistic (Steele and Torrie, 1960, p. 106) as the criterion measure. Pertinent information for the orthogonal comparisons is presented in Table 4. Least Significant Difference values were obtained for each comparison by multiplying the error variance by the square of each

TABLE 4

PERTINENT INFORMATION FOR THREE ORTHOGONAL COMPARISONS[a]

Item	Treatment groups						Least significant difference value	Comparison value
	ID-Same (A)	ID-New (B)	ID-Reversal (C)	ED-Same (D)	ED-New (E)	ED-Reversal (F)		
Mean log errors	.00000	.18573	.48083	.34082	.63543	.80069		
Error variance component	.00000	.24449	.24449	.24449	.24449	.24449		
Null hypothesis (H_0) and alternative (H_1) hypotheses for each comparison								
H_0[b]: A + B + C = D + E + F	+1	+1	+1	−1	−1	−1	−.844	−1.110[e]
H_1: A + B + C < D + E + F								
H_0[c]: A + D = B + C + E + F	+2	−1	−1	+2	−1	−1	−1.069	−1.421[e]
H_1: A + D < B + C + E + F								
H_0[d]: −A + B − C = +D − E + F	−1	+2	−1	+1	−2	+1	1.046	.239
H_1: −A + B − C ≠ +D − E + F								

[a] Orthogonal comparisons are defined by the coefficients entered in the treatment columns for each comparison. The sum of the coefficients of any row must equal zero and the sum of the cross products for any two rows must equal zero.

[b] Comparison between ID and ED-shift performance.

[c] Comparison of performance in the Same cue condition versus performance in the New and Reversal cue conditions.

[d] Comparison for interaction between dimension and cue conditions.

[e] $p < .01$.

coefficient and summing, and by using a tabular t value for 45 degrees of freedom. The findings were: (a) The average performance of the ID-shift groups was superior ($p < .01$) to the average performance of the ED-shift groups, (b) the average performance of the ID- and ED-*Same* groups was superior ($p < .01$) to the average performance of the ID- and ED-*New* and *Reversal* groups, and (c) there was no evidence of interaction ($p > .50$) between dimensional and cue conditions.

The average performance of the ID and ED groups in Stage 1 and Stage 2, and the average performance of each of the six groups in Stage 3 is plotted in Fig. 1.

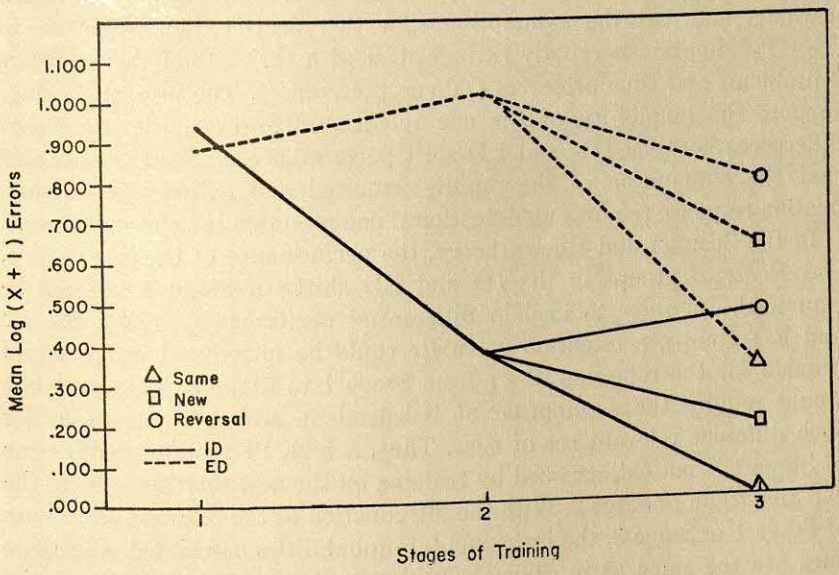


FIG. 1. The average performance of the three ID and three ED Groups in Stage 1 and Stage 2, and the average performance of each group in Stage 3.

DISCUSSION

The Zeaman and House (1963) prediction of superior ID-shift performance is confirmed by the comparison of ID and ED groups in Stage 3. Group comparisons of Stage 2 are not used to support the prediction of shift differences since different tasks were used in each shift. Shift differences are confounded with task difficulty; i.e., color is a more difficult dimension than form for retardates to learn (House and Zeaman, 1962). However, the ID and ED shifts of Stage 3 had identical transfer tasks and are free of the confounding effect of task difficulty. The use of identical shift tasks following different initial tasks is similar

to the transfer design recommended by Slamecka (1968, p. 425) for the evaluation of shift differences.

The Goodwin and Lawrence (1955) system accurately predicted the relative levels of performance in the cue conditions of Stage 3. Table 3 indicates that, within the ID and ED shifts, the performance of the *Same* groups was superior to the performance of the *New* and *Reversal* groups. It can also be seen in Fig. 1 that in both shifts the best performing groups were the *Same* groups and the groups having the most errors were the *Reversal* groups. According to Goodwin and Lawrence, the ability to predict performance in the cue conditions of Stage 3 requires: (a) The retention of cue preferences of Stage 1 across Stage 2 learning, and (b) the assumption that the identification responses to the OT-dimension is rapidly extinguished with the introduction of Stage 2 problems and thus preserves OT-cue preferences. The present findings support the possibility of the cue retention. However, the significant differences between ID- and ED-shift performance in Stage 3 indicated that the assumption of the rapidly acquired and extinguished identification response requires modification if cue retention is to be explained.

In the Zeaman and House theory, the performance of the *Same, New,* and *Reversal* groups in the ID and ED shifts of Stage 3 can not be explained. In order to explain differential performance among the ID and ED groups, a retention postulate could be introduced which would provide for the retention of Pr from Stage 1 to Stage 3. This postulate would require the assumption of independent probability systems for each different relevant set of cues. Thus, a high Pr for the positive cue of Stage 1 is not extinguished by training on the new relevant cues of the OT dimension in Stage 2. With the introduction of the relevant set of cues of Stage 1 in Stage 3, the instrumental probabilities associated with these cues are the same as in Stage 1. Within each set of relevant cues, the probabilities of approaching the positive and negative cues are mutually exclusive; i.e., Pr for the positive cue and 1-Pr for the negative cue.

REFERENCES

BERNSBERG, G. J., JR. Concept learning in mental defectives as a function of appropriate an dinappropriate "attention sets." *Journal of Educational Psychology,* 1958, **49**, 137–143.

CAMPIONE, J., HYMAN, L., AND ZEAMAN, D. Dimensional shifts and reversals in retardate discrimination learning. *Journal of Experimental Child Psychology,* 1965, **2**, 255–263.

GELLERMANN, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *Journal of Genetic Psychology,* 1933, **42**, 206–263.

GOODWIN, W. R., AND LAWRENCE, D. H. The functional independence of two discrimination habits associated with a constant stimulus situation. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 437–443.

HARLOW, H. F. Response by rhesus monkeys to stimuli having multiple sign values. *In* Q. McNemar and M. Merril (Eds.). *Studies in Personality.* New York: McGraw-Hill, 1942.

HOUSE, B. J., AND ZEAMAN, D. Reversal and nonreversal shifts in discrimination learning in retardates. *Journal of Experimental Psychology,* 1962, **63,** 444–451.

HOUSE, B. J., AND ZEAMAN, D. Miniature experiments in the discrimination learning of retardates. *In* N. R. Ellis (Ed.), *Child Development and Behavior.* New York: Academic Press, Vol. 1, 1963. (a)

HOUSE, B. J., AND ZEAMAN, D. Learning sets from minimum stimuli in retardates. *Journal of Comparative and Physiological Psychology,* 1963, **56,** 735–739. (b)

SLAMECKA, N. J. A methodological analysis of shift paradigms in human discrimination learning. *Psychological Bulletin,* 1968, **69,** 423–438.

STEELE, R. G., AND TORRIE, J. H. *Principles and Procedures of Statistics.* New York: McGraw-Hill, 1960.

ZEAMAN, D., AND HOUSE, B. J. The role of attention in retardate discrimination learning. *In* N. Ellis (Ed.), *Handbook of Mental Deficiency.* New York: McGraw-Hill, 1963.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 366–373 (1969)

Size-Constancy in Children Measured by a Functional Size-Discrimination Task[1]

JUDITH L. RAPOPORT[2]
National Institute of Mental Health

Since an instruction-produced attitude did not reliably affect the size-judgments of children below 10 years of age in a previous size-constancy task, a discrimination task for size at a distance was devised to compare the performance of three groups of children below the age of 10. With level of difficulty held constant for the three groups, there was no difference in the extent to which distance affected their size-discriminations. A college-age group did differ from the children, however. There was no simple relation between performance of the discrimination task and explicit attitudes about size-distance relationships, but the college-age group was better at size-discrimination, *per se*, and may have been attending to different aspects of the stimulus situation.

There was no significant correlation for individual differences between the functional size-discrimination task and a more traditional size-constancy task.

The majority of developmental studies of size-constancy have demonstrated an increase in constancy with age, but some investigators have not (Cohen, 1958; Frank, 1926). One problem in measuring constancy in young children is the difficulty in equating motivation, interest, and task attitude across different age groups. The results of some studies showing an age trend were probably biased by the need for considerable selection in those respects at the youngest ages (Beyrl, 1926; Piaget and Lambercier, 1951). Recent work has supported the idea that task attitude is an important factor in size-constancy judgments (Carlson and Tassone, 1967) and this finding has been extended to a school-age population above the age of 10 (Rapoport, 1967). Below age 10, task attitude could not be manipulated verbally and thus the wording of a given set of instructions cannot be presumed to insure a homogeneous approach to

[1] This research was supported by Public Health Service Special Fellowship 15,590. The author wishes to thank the Administration of the Alexander School, Silver Spring, Maryland, for their most generous cooperation in making space and subjects available for these experiments, Miss Sadie Fishman, for her skillful assistance in testing the children, and Mr. William H. Jones, for technical assistance.

[2] Laboratory of Psychology, National Institute of Mental Health, Bethesda, Maryland 20014.

the task across ages. Above 10, an age trend for size-constancy was obtained with one set of instructions, stressing objective physical size, while with another, emphasizing apparent size, it was not.

The present experiment represents an attempt to establish motivational homogeneity in a size-constancy situation, without having to rely on verbal instructions. Thus children's size-constancy was measured in a size-discrimination task in which functional discriminative capacity was assessed without invoking explicit judgments about size-distance relationships. It was hypothesized that with difficulty, motivation, and task attitude kept constant, there would not be an age trend for size-constancy. To determine whether performance on the discrimination task was in fact unrelated to a verbal concept of the size-distance relationship, a perspective task appropriate for use with young children was also devised. In this task, the children were asked how a stimulus object might be made to look farther away without actually moving it.

METHOD

Subjects. The subjects were 48 students at a nearby day camp and 16 college-age counsellors. There were four age groups (mean ages 5.4, 7.4, 10.1, and 20.3 years) with 8 males and 8 females in each group. The children were volunteers from the day camp who were told that they would help *E* try out a new game. The college-age subjects were told that they were a control group for the study of size-discrimination in children. An additional 32 paid college volunteers (Adult II in Table 1 below) were tested at a later date with the same procedure and apparatus in another setting.

Apparatus. The subject was seated at the head of a table 24 feet long, 24 inches wide, and 30 inches high. This was covered with a continuous piece of green felt upon which were five parallel HO-gauge electric train tracks placed 5 inches apart from each other, which ran the length of the table. On each track was a black HO-gauge electric train engine with one freight car attached. The stimulus material, five small black cardboard triangles were supported by black wooden platforms on the roofs of each of the five freight cars. The triangles were isosceles with an altitude-to-base ratio of 1. Each triangle differed in size. The smallest had a 3-inch height and there were four possible sets of triangles increasing in increments of $\frac{1}{4}$, $\frac{1}{8}$, $\frac{3}{32}$, or $\frac{1}{16}$ inch in base and height. The supporting blocks were all 1 inch square. The background was a white sheet of paper on the wall behind the far end of the table.

All five tracks were wired such that a train could be positioned automatically by *E* at any one of six possible places; a "home" position 1

foot 8 inches from the subject's end of the table, and positions 1–5 which were from 5 feet 5 inches to 20 feet 5 inches from the table end.

E sat at the side of the table near the subject. Directly in front of *S* was a control panel on which were five switches, each aligned with a train track and controlling the train for that particular track.

The testing was conducted in a room 30 × 12 feet. It was well illuminated with overhead fluorescent lighting. There was miscellaneous furniture in the room and cues to distance were good.

Procedure. Subjects were asked to select the largest (or smallest) triangles when the trains were all at the same distance (either 1 foot 8 inches, or 20 feet 5 inches from *S*). Instructions were: "We are going to play a game to see if you can tell which of these is the biggest (littlest) triangle. Choose the right one by calling the train home with the switch in front of its track. Can you pick the biggest one now?"

After the *S* had chosen, he was told if he was correct and asked to pick the biggest (littlest) of the four remaining triangles until all but one was chosen. Thus, for each stimulus display of five triangles, *S* made four choices.

The experimenter altered the sets of triangles until *S*s made no more than one error in the choices for both distances combined. This procedure had been found to insure an error leveled about 50% when the stimuli were discriminated at different distances from each other. For most children, ⅛ inch step sizes between triangles were used, for the adults, 3/32 inch was most frequent.

The following instructions were given: "We will now make the game harder by placing each train in a different place along the table. You must continue to do as you have done, calling home the biggest (littlest) first, and then the biggest (littlest) of those that are left on the table, until the trains are all called in."

After each choice, the subject was told if he had chosen correctly or not and was encouraged to do well. Between trials, the subjects were given other tasks (described below) while *E* repositioned the triangles on the trains, and the trains on their respective tracks, according to a random schedule such that each triangle appeared equally often on any track and in any of the five positions. No two trains were ever at the same distance during an experimental trial.

Half of the subjects in each age group were instructed to choose the largest triangle and half the smallest. As a measure of whether attention had been sustained over the testing period the *S*s were retested at the end of the session with the triangles all at the same distance.

Between trials subjects were given other experimental tasks in another room where they could not see the trains being repositioned for the next

trial. Upon completion of the task, subjects were asked for a qualitative description of how they had gone about the task. The seated eye height of each *S* was measured.

Additional Tests

Size-constancy. Eight *Ss* from each of the day camp groups and all of the second adult group were given a standard size-constancy task in which a far variable triangle was matched with a near standard. The procedure was similar to that of Carlson (1967). Two instructions, Objective and Apparent size were given each *S*. The results for this task are part of a larger study discussed more extensively elsewhere (Rapoport, 1967).

Perspective task. The three children's groups were shown the apparatus for the size-constancy task described above with both triangles at the same distance of 5 feet from the *S*. The *Ss* were asked, "How can we change the size of this triangle so that it looks twice as far away as the other one without really moving it?" This was repeated five times for each *S*.

Results

The experimental task was highly successful in engaging the interest and attention of the children. Two 5-year-old subjects had to be discarded because they could not reliably make the size discrimination when the triangles were all at the same distance with the easiest set of triangles. For most of the children, ⅛ inch-step sizes between triangles could be discriminated perfectly when at the same distance from *S*, but gave approximately a 50% error level when the trains were each at different distances in the experimental trials. This error level provided sufficient errors to handle statistically, while ensuring enough success to keep the children motivated.

Attention of all groups was sustained throughout the testing period as measured by their ability to discriminate the sizes at the same distance at the end of the session, as well as by the lack of any consistent trend in error level from the beginning to the end of the test.

As each subject was presented the experimental display of trains eight times, he made a total of 32 choices (four choices for each stimulus array of five trains). The results are given for each age group in Table 1. Error Level (EL) refers to the total percentage of the 32 responses which were incorrect, i.e., the triangle chosen on that choice was not the largest (smallest) of those displayed. Projective Errors (PE) are the percent of those errors which followed a retinal angle match, that is, the *S*'s

TABLE 1
MEAN PROJECTIVE RESPONSES ON SIZE-DISCRIMINATION TASK

	Age group				
	5	7	9	Adults I	Adults II
Percentage of projective errors	76	75	81	56	59
SD	24	18	19	35	23
Error level (%)	47	46	47	38	39
N	16	16	16	16	32

choices were too near when the task was to choose the largest triangle and too far when instructions were to choose the smallest triangle.

The absolute error level was almost identical for the three children's groups but was clearly less for both adult groups than for the children. Most of the children used the ⅛ inch-step size between triangles, although a few of the youngest group had to be given the easier set (¼-inch steps) and several of the 9-year-olds required 3/32-inch steps to produce a single error in the initial discrimination task. The adult group, however, used either the 3/32 or 1/16-inch sets, as they could discriminate the sizes when the triangles were at the same distance better than the children could. Even with this more difficult task a 50% error level could not be obtained. Within each group there was no correlation between step size used, error level, eyeheight, and percentage of projective errors.

Percentage of projective errors, the measure of the relation of distance to errors in the size-judgments, did not differ significantly for the three children's groups ($F = 0.410$, $df = 2,36$). All children's groups showed relative underconstancy in the sense that their errors were greater than 50% in relation to projective size. These results were highly consistent. There were only 4 individual exceptions out of 48 children. However, the adult groups showed no relation of errors to projective size and when they are included in an analysis of variance, there is a significant change with age for percentage of PE ($F = 2.94$, $p < .05$, $df = 3,48$).

There was no significant correlation between the percent of projective errors on the train-task and the size-constancy measure by triangle setting under either instruction, for any of the children's or adult groups.

The adult group gave qualitatively different descriptions of how they made their size-judgments. Some stated that they attended to the ratios of each triangle to its supporting block while others compared the triangles as they passed each other when they were brought back by *S* in an active way. Children did not verbalize as much activity and more uniformly indicated that they simply picked one which "looked" largest or smallest.

Perspective task. If a child set the triangle smaller, three or more out of the five times, in response to the question, "Make this triangle look twice as far away," it was scored as a "perspective" response.

As Table 2 indicates, there was a clear progression over this age range toward a perspective setting for this task.

TABLE 2

TRIANGLE SETTING FOR PERSPECTIVE TASK[a]

Age	Set smaller	Set bigger	Unclear
5	7	7	2
7	10	4	2
9	15	1	0

[a] Response to task of making triangle look "far away."

DISCUSSION

One striking feature of these results is the high accuracy of size-at-a-distance judgments of all groups with this task. Even the youngest group made ⅛-inch size discriminations over a 21-foot distance with greater-than-chance frequency.

The hypothesis of no age-trend for size-constancy was confirmed in the respect that there was no difference among the children's groups in the percentage of projective errors on the discrimination task. Similarly, no support for the influence of an explicit concept of a perspective size-distance relationship on the experimental task was found; the 5 and 9-year-old groups had almost identical percentage of Projective Errors although differing greatly in their responses to the perspective task. In addition, there was no correlation for the children's or adults responses in the size-discrimination task with performance on the standard size-constancy measure under either "Objective Size" or "Apparent Size" instructions.

In spite of the absence of an age trend within the children's groups, there was a significant difference between children's and adult groups in the percentage of Projective Errors—adults errors were not significantly related to distance while the children's were. Adults were markedly better at the absolute size-discrimination task (with the triangles at the same distance); with the most difficult set of triangles, they still made fewer errors than the children. It may be that the adult judgments were so nearly perfect that a qualitatively different task was performed in which the sources of error were different from those for the children. Subjective descriptive reports indicated that the adult group was more likely to actively consider the relation of each triangle to its immediate surround, i.e., the supporting wooden block. None of the children gave

such an account which may reflect the fact that when the discrimination was difficult they were more likely to project the triangles to the background. It is possible that these differing accounts reflect just a difference in verbal ability. A difference between adults and children in the strategy of size judgment might be tested experimentally by varying the size of the supporting blocks and obtaining a greater effect on the adults than on the children's responses. As the triangles were supported on the train roofs, it was easier to project them to a common background than it would have been if they rested on the tabletop. This finding is similar to that of Cohen *et al.* (1958) in which an age trend in a size-constancy task was found only with their more difficult discrimination task, over a comparable age-range.

The results presented here are at variance with the recent study of Leibowitz *et al.* (1967), which compared the size-constancy of children and college students over a comparable age range and showed progressively increasing size-constancy over the entire age span. The argument made by Leibowitz *et al.* (1967) is that, as the results were obtained with unfamiliar objects, perceptual rather than cognitive cues are the important determinants of the relative degree of size-constancy. However, the Leibowitz study utilized objective instructions which have been demonstrated to bring out a bias toward greater size-constancy in adults (Carlson and Tassone, 1967). Moreover, this bias increases with age (Rapoport, 1967). In addition, the groups over 10 years of age in the Leibowitz study produced size matches that were actually in the range of overconstancy, and it is easier to explain overconstancy on a cognitive than on a perceptual basis. Also, as infants have been shown to manifest considerable size-constancy (Bower, 1965), it seems unlikely that perceptual learning, in the sense of a gradually decreasing reliance on the retinal angle, takes place between ages 10 and 20.

Since the basic development of size-constancy as a perceptual capacity would seem to have occurred before the age of 5 years, and the conceptual representation of distance by decreased object size to occur thereafter, this problem may be an instance where the development of a cognitive ability depends upon the prior development or existence of a related underlying perceptual capacity. A second important aspect of these findings is the lack of correlation between individual size-constancy measures with the discrimination task and those obtained with the more classical triangle setting, although comparable distances were used. There has been considerable theorizing about the significance of individual differences on size-constancy measures, but the lack of a significant correlation between the measures under two different techniques casts considerable doubt on the generalizability of such findings.

REFERENCES

Bentz, F. Über die Größenauffassung bei Kindern. *Zeitschrift für Psychologie*, 1930, 116, 344–371.

Bower, T. G. R. Discrimination of depth in premotor infants. *Psychonomic Science*, 1965, 1, 368.

Carlson, V. R., and Tarrow, E. P. Independent size-judgments at different distances. *Journal of Experimental Psychology*, 1967, 73, 491–497.

Cohen, W., Hirshkowitz, A., and Chisholk, M. Size judgment at different distances as a function of age level. *Child Development*, 1958, 29, 473–479.

Frank, H. Untersuchung über Subgrößenkonstanz bei Kindern. *Psychologische Forschung*, 1926, 7, 127–145.

Leibowitz, H., Pollard, S. W., and Dickman, D. Monocular and binocular size-matching as a function of distance at various age levels. *American Journal of Psychology*, 1967, 80, 263–268.

Piaget, J., and Lambercier, M. Recherches sur le développement des perceptions VIII. Transpositions perceptives et transmitté opératoire dans les comparaisons en profondeur. *Archives de Psychologie*, Geneve, 1946, 31, 225–308.

Piaget, J., and Lambercier, M. Recherches sur le développement des perceptions XII. La comparaison des grandeurs projective chez l'enfant et chez l'adulte. *Archives de Psychologie*, Geneve, 1951, 33, 81–130.

Piaget, J., and Lambercier, M. Recherches sur le développement des perceptions XXIX. Grandeurs projectives et grandeurs réelles avec étalon éloigné. *Archives de Psychologie*, Geneve, 1956, 35, 257–286.

Rapoport, J. L. Attitude and size judgment in school-age children. *Child Development*, 1967, 38, 1183–1192.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 374–384 (1969)

Dimension Preference and Performance on a Series of Concept Identification Tasks in Kindergarten, First-Grade, and Third-Grade Children[1]

MERRILL M. MITLER AND LAUREN HARRIS
Michigan State University

Data are presented relating children's preferences for stimulus dimension (form, color, and number) to performance on concept-identification tasks involving preferred and nonpreferred dimensions. The *Ss* were 77 kindergarten, first-, and third-grade children. As the measure of dimension preference, all *Ss* sorted cards from the Wisconsin Card Sorting Test. They then performed concept-identification tasks employing the same cards. Seventy-five of the 77 *Ss* showed significant dimension preferences, 59 *Ss* preferring form, 9 number, and 7 color. All 59 form-preferring *Ss* eventually reached criterion for each dimension task. Trials to criterion on each task, however, varied significantly among levels in school and among dimensions. Performance on the concept identification tasks was directly related to level in school. Across levels in school, the form task was most easily solved, color next, and number least.

The superiority of third-graders to kindergarteners on the concept identification tasks was apparently not a function of immediate prior experience on the preference test, since the same difference obtained in control kindergarteners and third-graders who received the concept tasks *before* the preference test. On the number task, however, significantly more control than experimental kindergarteners reached criterion. Furthermore, 47% of the control *Ss* had significant dimension preferences compared with 97.4% of the experimental *Ss*. These differences apparently resulted from differences in immediate prior experience.

There is much research showing developmental changes in how children group, or categorize, multidimensional objects in free-sort situations. As early as 1914 Descoudres found that with increasing age, preference for color decreased, and preference for number increased. Preference for form remained relatively high and constant with increasing age. Des-

[1]This paper is based on a thesis submitted by the first author, under the second author's direction, to the Department of Psychology, Michigan State University, in partial fulfillment of the MA degree requirements. The research was supported by a grant to the second author from Michigan State University. Responsibility for design, execution, and report of this research rests equally with both authors. Portions of this research were reported at the meetings of the Midwest Psychological Association, Chicago, 3 May, 1968.

coudres held that her dimension preference findings could not be construed as indexes of whether other dimensions were perceived, since many *Ss* asked questions specifically naming the other dimensions involved. According to Descoudres, these questions indicated that all dimensions were "available" for at least some *Ss*.

Later studies have dealt primarily with color and form preferences and, generally, have corroborated Descoudres' findings. Colby and Robertson (1942) found a decrease in frequency of color preference in children ranging in age from 3:6 to 9 years. Kagan and Lemkin (1961) enlarged the scope of the question by including size as a dimension. They found that preference for form was greatest, color next, and size least. Brian and Goodenough (1929), however, reported an initial preference for form in their youngest *Ss* (22 months old) followed by a preference for color in their 3- to 6-year-old *Ss* and finally a preference for form again in their older *Ss* (6 to 14 years old). Corah and Gross (1967) were among the first investigators to study dimension preference through manipulation of stimulus variables. They found that reducing the saliency of the color dimension increased the proportion of form choices in kindergarten-age *Ss*. Corah's work and other findings (Huang, 1945) thus suggest that very young children match stimuli on the basis of the dimension along which the differences between stimuli are most discriminable.

Recently Suchman and Trabasso (1966) extended this kind of work in an important way: by relating the child's dimension preference in a free-sort situation to his performance on concept identification tasks that involved the same stimulus dimensions. They found that concept tasks along the preferred dimension were solved more easily, suggesting that dimension preference both sets the initial sorting response and influences the order of dimensional concept attainment.

Suchman and Trabasso studied the dimensions of color and form in nursery- and kindergarten-age children. The present study expanded the scope of their study by including the dimension of number and by testing a wider grade range (kindergarten through third grade).

METHOD

Subjects. The *Ss* were 38 girls and 39 boys. The sample included 36 kindergarteners (5 years 3 months–6 years 7 months), 20 first-graders (5:4–7:1), and 21 third-graders (8:4–9:2). There was approximately the same number of girls and boys in each grade.[2]

Materials. The stimulus and response cards were selected from the

[2] We are grateful to Principal Arthur E. Lange, the staff, and the children of the Spartan Village Public School, East Lansing, Michigan for their enthusiastic cooperation.

Wisconsin Card Sorting Test (Berg, 1948). Each card is a 3 × 3-inch square. Four sets of stimulus cards were used. A set consisted of four cards: one card displayed 1 red triangle; another displayed 2 green stars; another, 3 yellow crosses; and the last, 4 blue circles. Each set of stimulus cards was mounted on a 24 × 5-inch Masonite board, a different order for each set. The orders were 1234, 4321, 2413, and 3142. The stimulus cards were mounted so that when the boards lay with the long edge down, the stimulus cards were upright and in one of the above orders from left to right.

The deck of response cards contained 24 of the 64 response cards in the Wisconsin Card Sorting Test. To eliminate ambiguity on the test of dimension preference as well as the partial reinforcement effect on the concept-identification tasks that followed, these 24 cards were selected so that on any one sort, *S* could not use the same stimulus card for more than one dimension. For example, the response card displaying 3 yellow circles was omitted because the stimulus card displaying 3 yellow crosses could be used to sort both for color and for number.

The stimulus configurations on all cards varied from card to card. The order of the response cards was random with the restriction that no one of the four levels in each of the three dimensions was repeated more than twice in succession. Thus a series of the same color, form, or number never exceeded two in length.

Procedure. Each *S* accompanied *E* to a spare room in the school for individual testing. The *S* sat at a desk facing a wall. The *E* and a second adult sat on *S*'s right and left sides, respectively. The second adult recorded *S*'s responses on score sheets correlated to the particular sequence of tasks for each *S*. The stack of four stimulus boards was propped against the wall facing *S*. One board lay behind the other so that only one stimulus card order was visible at a time. The order of board presentation was varied across *S*s using the same orders as for the stimulus cards.

The *E* spoke the following instructions from memory:

> "See these cards up here? Take each one of these cards and put it up here in front of the one you think it goes with. Put each of these in front of one up here, the one it goes with. Try the first one. I'll take each away after you put it down so there will be room for another. Go ahead."

The *E* aided his instructions with gestures and pointing. For each trial, *E* handed the response card to *S* making sure the card was in upright orientation.[3]

The *S* placed each card flat on the table directly under the chosen

[3] Crucial for the 5-pointed stars, crosses, and triangles.

stimulus card. With every six responses, the recorder flipped down the visible stimulus board revealing the next board. This was the procedure for the first part of the experiment and is referred to as the Dimension Preference Test (DPT).

Once *S* completed the 24 sorts in the DPT, the recorder set the stimulus boards to the same ordering as before the DPT. Then *E* spoke the following instructions from memory, gesturing as before:

> "Now let's try something different. You take each of these cards. For each one of these, I'll pick one from up here, one I think it goes with. Only I'm not going to tell you which one it is. You have to guess. Put each one of these cards in front of the one I'm thinking of. I'll tell you if you guess right after each one. You try to guess right as many times as you can. Try the first one."

The *E* presented each response card as before, this time using cards from a duplicate response deck. This deck was arranged in a different order with the same restrictions as before. The *S* placed each card as he had in the DPT. After each response, *E* said, "Yep, that's the one," or "Nope, that's not it." For each *S* the sequence of board presentation was recorded to insure a record of stimulus card position throughout the session. Every six trials, as before, the recorder flipped to the next stimulus board. After 24 trials, all stimulus boards were put in their initial position. The 25th and the 49th trials proceeded as did the first.

Each *S* was allowed a total of 72 trials to complete a series of six dimensional problems. The three series of six concept identification tasks used for this part of the study were: 1. FCNCFN (Form, Color, Number, Color, Form, Number); 2. CFNFCN, 3. NCFNFC. For each *S* a series of tasks was preassigned. The *E* said "Yep" only after correct sorts. After six consecutive correct sorts on one dimension, *E* said "Yep" to sorts on the following dimension in *S*'s series of tasks. The session ended after the 72nd trial regardless of *S*'s progress in the series. Thus the Ss assigned to the CFNFCN series had to make six consecutive correct sorts on color, six on form, and so on to complete the series. The only information given other than instructions was *E*'s saying "Yep" or "Nope" after each trial.[4]

RESULTS AND DISCUSSION

Dimension preference test. Calculated by binomial expansion ($p = 1/3$ for sorts along dimension under analysis, $q = 2/3$ for sorts along other

[4] Pilot work revealed that *E*'s saying "yes" and "no" or "right" and "wrong" seemed to make many children tense and unwilling to complete the session. "Yep" and "nope" seemed to make the tasks more game-like.

dimensions), the criterion for significant preference for one of the three dimensions on the DPT required that *S* make a total of at least 16 of his 24 sort on one dimension. If sorts on each of the three dimensions were equally likely, the probability of 16 sorts on one dimension is less than .003. With this criterion, 75 of the 77 *Ss* showed significant preferences; of these 75 *Ss*, only 13 made fewer than all 24 sorts on one dimension. There seemed to be no systematic pattern in the distribution of these 13 children over the levels in school.

Thus the frequency of preference was very high. Furthermore, if number of sorts along a dimension reflects preference for that dimension, the over-all tendency was for strong preferences, and strength of preference was not related to level in school.

Table 1 summarizes the results for the DPT. It can be seen that form

TABLE 1
RESULTS OF THE DIMENSION PREFERENCE TEST AS A FUNCTION OF LEVEL IN SCHOOL AND SEX

	Level in school						Total
	Kindergarten		First grade		Third grade		
Dimension	Boys $N = 13$	Girls $N = 23$	Boys $N = 11$	Girls $N = 9$	Boys $N = 15$	Girls $N = 6$	$N = 77$
Color							
Number of *Ss*	2	3	0	0	2	0	7
	15.4%	13.0%	0%	0%	13.4%	0%	9.1%
Form							
Number of *Ss*	11	16	10	8	9	5	59
	84.6%	69.5%	91.0%	89.0%	60.0%	83.5%	76.6%
Number							
Number of *Ss*	0	3	1	1	3	1	9
	0%	13.0%	9.0%	11.0%	20.0%	16.5%	11.7%
No preference							
Number of *Ss*	0	1	0	0	1	0	2
	0%	4.5%	0%	0%	6.6%	0%	2.6%

was generally the preferred dimension, and the data suggest no relation between preference for other dimensions and level in school. Table 1 also reveals the absence of sex differences in preference for color, form, and number dimensions.

Concept identification tasks. The means for number of tasks completed among the three grade levels were: kindergartners, $\bar{X} = 1.78$ (SD = 1.22); first-graders, $\bar{X} = 2.75$ (SD = 1.67); third graders, $\bar{X} = 3.33$ (SD = 1.79). Analysis of variance revealed a significant effect of grade ($F = 7.2$, $df = 2/74$, $p < .01$). Results of cell contrasts indicated that

kindergarteners differed significantly from first-graders ($F = 5.92$, $df = 1/54$, $p < .05$) and from third-graders ($F = 14.5$, $df = 1/55$, $p < .001$). First-graders did not differ significantly from third-graders, however ($F = 1.2$, $df = 1/39$, $p > .05$).

Total errors on the 72-trial session also reflect performance on the task series. The mean number of errors for the three grade levels were: kindergarteners, $\bar{X} = 47.67$ (SD = 9.34); first-graders, $\bar{X} = 42.30$ (SD = 10.08); third-graders, $\bar{X} = 35.00$ (SD = 11.63). Analysis of variance again revealed a significant effect of grade ($F = 10.9$, $df = 2/74$, $p < .01$). Results of cell contrasts disclosed that the only significant difference was between kindergartners and third-graders ($F = 14.5$, $df = 1/55$, $p < .001$).

Dimension preference and performance on the concept identification tasks. To establish the relation between dimension preference and performance on the concept identification tasks required comparison among levels in school and preferences with respect to the number of trials to criterion for the first concept task. Since the ordering of tasks was varied among *Ss*, some *Ss* had the color concept first, some had form, and others had number. Thus a 3×3 factorial design, level in school *versus* dimension of first task, could be constructed for form-preferring *Ss*, another 3×3 design for color-preferring *Ss*, and a third for number-preferring *Ss*. Table 2 presents the average number of trials to criterion on the first task for all of the 75 *Ss* with significant preferences. Unfortunately the frequencies of color and number preferences were too small to permit planned statistical treatments.

Considering only the form preferrers (3 columns at left of table), a two-way analysis of variance, level in school by dimension of first task, indicated that the number of trials to criterion was inversely related to level in school ($F = 3.98$, $df = 2/50$, $p < .05$). Performance of the form-preferrers also was related to the dimension of the first task ($F = 7.52$, $df = 2/50$, $p < .01$) such that these *Ss* did better when form was the first task. (Note that performance scores of the color and number preferrers also appear to be related to the preferred dimension and to the dimension on the first task.)

As can be seen from Table 2, the difference between kindergarten and other *Ss* for the number task was the major source of variance. However, the level-in-school by dimension interaction was not significant ($F = 2.10$ $df = 4/50$ $p > .10$). There were no significant differences between kindergarteners and other *Ss* for the color or form tasks (Fs ≤ 1.9, $df = 2/16$, $p > .30$).[5]

[5] Although differences among the means for color appear quite large, they in fact result from one kindergartener who scored disproportionately high on the color task as reflected in the standard deviation in that cell.

TABLE 2

MEAN NUMBER OF TRIALS TO CRITERION ON FIRST TASK AS A FUNCTION OF DIMENSION INVOLVED IN FIRST TASK FOR ALL 75 *Ss* WITH SIGNIFICANT DIMENSION PREFERENCES

Dimension on first task	*S*'s preferred dimension on DPT								
	Form			Color			Number		
	Form	Color	Number	Form	Color	Number	Form	Color	Number
Kindergarten									
N	10	8	9	0	2	3	1	2	0
$\bar{X}$	7.0	26.4	57.8		6.0	65.0	31.0	22.5	
SD	3.16	27.25	25.96		0	9.90		22.47	
First-grade									
N	7	7	4	0	0	0	1	0	1
$\bar{X}$	17.0	12.8	28.3				11.0		6.0
SD	24.36	6.10	30.10						
Third-grade									
N	5	5	4	0	1	1	1	1	2
$\bar{X}$	8.0	9.6	18.8		6.0	6.0	7.0	9.0	13.0
SD	5.50	3.00	20.87						7.00

At this point, the question arose whether the difference on the number task reflected a true difference in ability to identify number or whether *Ss* had been "set" or predisposed to sort for *form* by their immediately preceding experience with the DPT and that kindergarteners were more susceptible than third-graders to such a set. We hoped to answer this question by assessing performance without the influence of prior experience with the DPT.

Therefore, 17 control *Ss* (11 kindergarteners and 6 third-graders) were tested on the concept task series *before* they were given the DPT. Because the greatest difference in the experimental *Ss* was between kindergarteners and other *Ss* on the number task, we made number the first task for the control *Ss*.

Table 3 compares the control *Ss*' scores with those of their experimental counterparts on the number dimension. Proportionately more control kindergarteners (7 of 11) reached criterion than did experimental kindergarteners (3 of 9; $t = 1.76$, $df = 19$, $p < .05$, one tailed[6]).

However, the relative difference between kindergarteners and third-graders in the control group remained significant ($F = 6.1$, $df = 1/16$, $p < .01$). For third-graders, all experimental and control *Ss* reached criterion on the number task. Mean number of trials to criterion there-

[6] Since the hypothesized "setting" effect of the DPT was to retard performance on the number task, a one-tailed t test of these frequencies was deemed justified.

TABLE 3

MEAN NUMBER OF TRIALS TO CRITERION FOR FIRST DIMENSION (NUMBER) FOR CONTROL *Ss* AND COMPARABLE EXPERIMENTAL *Ss*

Level in school	Mean	
	Control *Ss*	Experimental *Ss*
Kindergarten	$N = 11$	$N = 9$
	(SD = 26.87)	(SD = 24.55)
	9.2	18.8[a]
Third-grade	$N = 6$	$N = 4$
	(SD = 2.67)	(SD = 16.4)
Total	$N = 17$	$N = 13$
	37.2	57.8

[a] Because there were only 4 *Ss* in this experimental group-cell, 4 additional *Ss* were tested. A one-way analysis still failed to disclose a difference between control and experimental third-graders on the number task. (Experimental mean = 23.9, $N = 8$, Control mean = 9.2, $N = 6$, $F = 1.48$, $df = 1/12$, $p > .40$.) It should be mentioned that the means of 18.8 and 23.9 are distorted by two high scores and therefore are not representative.

fore was substituted as the comparison statistic. The differences were not significant ($t < 1.0$).

We may conclude that prior experience may differentially affect performance on concept identification tasks as a function of the age of the child, since the control kindergarteners performed significantly better on the number task than did their experimental counterparts, while the control and experimental third-graders were not significantly different.

Experimental and control performance on the DPT. Of the 77 experimental *Ss*, 75 (97.4%) had significant dimension preferences compared with only 8 (49%) of the 17 control *Ss*. The difference in proportions was significant ($t = 7.6$, $df = 93$, $p < .001$). The 8 control *Ss* with significant preferences comprised two kindergarteners and 1 third-grader who preferred number, 2 third-graders who preferred color, and 3 kindergarteners who preferred form. The performance of the control children on the DPT thus suggests that the concept identification tasks themselves can have setting effects for dimension preference when they immediately precede the preference test.

In summary, these results indicate the considerable importance of immediate prior experience in the assessment of both dimension preference and ability to solve multidimensional concept-identification problems. Future research on the relation between dimension preference and concept identification should be designed to allow for the possibility of mutual contamination of preference and performance measures.

More generally, the results on the Dimension Preference Test are consistent with other data insofar as form is usually the preferred dimension after 5 years of age (Descoudres, 1914; Kagan and Lemkin, 1961). These data, however, disagree with those of Brian and Goodenough (1929), who found strong preference for color from 3 to 6 years of age preceded and followed by periods of form preference. This finding has puzzled workers in the area for many years because of the seeming turnabout in the developmental sequence that cognitive theorists, such as Heidbreder (1946), would predict. Perhaps the difference between their results and the current findings stems from genuine differences between the pace of children's perceptual development in the 1920's and today. But this interpretation seems forced in light of Descoudres' findings which appeared just 15 years earlier than Brian and Goodenough's. The difference might lie in method: Brian and Goodenough tested their over-5-year-old *S*s only with planometric forms. But children under 5 years were given stereometric forms first and permitted to handle them, after which they were tested on the planometric forms. This procedure could have constituted a pretraining or sensitization condition which emphasized the dimension of shape. It is possible that in Brian and Goodenough's under-5-year-old group, the younger children—those under 3—were more susceptible than were the older children to such pretraining experiences. The present data would support this suggested relation between age and susceptibility to pretraining. Furthermore, Brian and Goodenough excluded from their analyses many of the children under 3 years of age whose responses, in their estimation, indicated little understanding of the task. It is possible, then, that Brian and Goodenough's procedure artifactually produced form matches among the children under 3 years of age, the finding of early form preference being a product of subject sampling, differential susceptibility to pretraining, and different experimental treatment.

Dimension availability, preference, and usability. The results for the experimental *S*s suggest to us that there are at least three logically different factors involved in a child's conceptual treatment of stimulus dimensionality. The first two factors are dimension "availability" and dimension "preference." A particular dimension is available for an individual if he can respond to two objects' mutual similarity and to their qualitative difference from a third object along that dimension. The second factor is dimension "preference"; a child given a free choice might sort—for whatever reason—exclusively or primarily along one available dimension. These two factors, however, cannot be distinguished from one another using only the multidimensional free-sort situation. For ex-

ample, on the DPT most of the experimental *Ss* sorted a significant number of times on the form dimension. These children may be said to have "preferred" that dimension. By logical implication, it follows that the form dimension was "available" to them. Nothing, however, can be said yet about availability and preference of color or number for these children. But on the concept-identification tasks all the children succeeded in reaching criterion on the color task in addition to the form task. Color, therefore, was shown to be an available dimension while not being a preferred dimension.

Neither availability nor preference by themselves, however, can logically account for success in a concept-identification task involving a particular dimension. We also found that among the subjects who did solve all three tasks, trials to criterion differed significantly as a function of dimension. Thus those children who reached criterion on the number task did so less readily than on the form task. To this extent, the number dimension, while available (and not a preferred dimension), was less "usable" as a cue in a learning situation. Our third factor, then, is dimension "usability." By definition, availability is a dichotomous construct: a dimension is either available or not available. Then, the extent to which stimulus differences along an available dimension can function as cues in a concept-identification situation is the extent to which the dimension is *usable* for the child. Our evidence suggests that for form-preferring children, across the levels in school, form was most usable, color next, and number least usable as measured by number of trials to criterion on the concept-identification tasks. Since some kindergarteners on the concept-identification tasks made number responses at no more than chance levels, let alone reached criterion, number may not even have been available for these children. (Such children, needless to say, did not prefer number on the dimension-preference test.) Among levels in school and across dimensions, kindergarteners were least able to use the dimensions, and third-graders, the most.

The phenomenon of dimension preference clearly is more complex than the literature to date has indicated. The three factors, dimension availability, preference, and usability, have, perhaps, been confused in past research where, for instance, absence of a preference for a particular dimension in a free-sort situation is taken as evidence for the nonavailability of that dimension. These factors should be mutually distinguished in future research on conceptual development, indeed in all kinds of discrimination-learning situations, concept identification tasks being merely a subset. Because they lie closer to the data language these factors should be helpful in structuring new experimental designs as well.

REFERENCES

BERG, E. A. A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, 1948, **39**, 15–22.

BRIAN, C. R., AND GOODENOUGH, F. L. The relative potency of color and form perception at various ages. *Journal of Experimental Psychology*, 1929, **12**, 197–213.

COLBY, M., AND ROBERTSON, J. Genetic studies in abstraction. *Journal of Comparative Psychology*, 1942, **33**, 385–401.

CORAH, N. L. Color and form in children's perceptual behavior. *Perceptual and Motor Skills*, 1964, **18**, 313–316.

CORAH, N. L., AND GROSS, J. B. Hue, brightness and saturation variables in color-form matching. *Child Development*, 1967, **38**, 137–142.

DESCOUDRES, A. *Couleur, forme, ou nombre? Archives de Psychologie*, 1914, **11**, 305–341.

HEIDBREDER, E. The attainment of concepts: II. The problem. *Journal of General Psychology*, 1946, **35**, 191–223.

HUANG, I. Abstractions of form and color in children as a function of the stimulus objects. *Journal of Genetic Psychology*, 1945, **66**, 59–62.

KAGAN, J., AND LEMKIN, J. Form, color, and size in children's conceptual behavior. *Child Development*, 1961, **32**, 25–28.

SUCHMAN, R. G., AND TRABASSO, T. Stimulus preference and cue function in young children's concept attainment. *Journal of Experimental Child Psychology*, 1966, **3**, 188–198.

Errata

Vol. 6, No. 2 (1968), in the article, "Response Variability in Retarded Children," by John J. Gates and Dean L. Fixsen, pp. 306–320:

Page 309: Footnote number 3 should read as follows:

The first is that the PI equals one only when one key received all responses. If more than one, but less than eight keys received responses, the least preferred key was taken as the one that received no responses. In other words, keys that received no responses were included in computing the range. The second limitation is that different response distributions *may* result in the same PI value, while their relative frequency graphs might indicate a higher preference in one case and a lower preference in the other. Relative frequency graphs are included in the figures to permit comparison of the two measures. It is felt that the economy of data presentation afforded by the PI warrants its use, even with the above qualifications.

Vol. 6, No. 4 (1968), in the article, "Age Changes in Perception of Verticality and of the Longitudinal Body Axis under Body Tilt," by Seymour Wapner, pp. 543–555:

Page 553, line 10 from bottom, " . . . perception progresses verticality." should read: " . . . perception progresses toward veridicality."

Journal of
APPLIED BEHAVIOR ANALYSIS

MONTROSE M. WOLF, Editor
DONALD M. BAER, Associate Editor
VICTOR G. LATIES, Executive Editor

EDITORIAL BOARD

T. Ayllon
N. H. Azrin
A. Bandura
W. C. Becker
J. S. Birnbrauer
C. B. Ferster
I. Goldiamond
J. G. Holland
F. S. Keller
P. Lang
O. R. Lindsley
O. I. Lovaas
J. L. Michael
G. R. Patterson
T. R. Risley
J. A. Sherman
M. Sidman
G. Siegel
B. F. Skinner
J. E. Spradlin
A. W. Staats

Second quarterly issue, SUMMER, 1968

Behavioral engineering: postural control by a portable operant apparatus.

N. Azrin, H. Rubin, F. O'Brien, T. Ayllon, and D. Roll.

Establishing use of descriptive adjectives in the spontaneous speech of disadvantaged preschool children.

Betty M. Hart and Todd R. Risley.

Effects of candy and social reinforcement, instructions, and reinforcement schedule leaning on the modification and maintenance of smiling.

B. L. Hopkins.

Feedback in behavior modification: an experimental analysis in two phobic cases.

Harold Leitenberg, W. Stewart Agras, Laurence E. Thomas, and Dale E. Wright.

Rules, praise, and ignoring: elements of elementary classroom control.

Charles H. Madsen, Jr., Wesley C. Becker, and Don R. Thomas.

An experimental analysis of verbal imitation in preschool children.

Thomas A. Brigham and James A. Sherman.

Operant conditioning of eye movements.

Stephen R. Schroeder and J. G. Holland.

Collateral social development accompanying reinforcement of outdoor play in a preschool child.

Joan Buell, Patricia Stoddard, Florence Harris, and Donald M. Baer.

A method to integrate descriptive and experimental field studies at the level of data and empirical concepts.

Sidney W. Bijou, Robert F. Peterson, and Marion H. Ault.

Address subscription orders to Mary L. Sherman,
Department of Human Development, University of Kansas
Lawrence, Kansas 66044

Students $4 * Individuals (personal use) $8 * Institutions $16

The Psychology of Learning and Motivation

Edited by the late **Kenneth W. Spence**
and **Janet Taylor Spence**
Department of Psychology
University of Texas, Austin, Texas

Volume 1

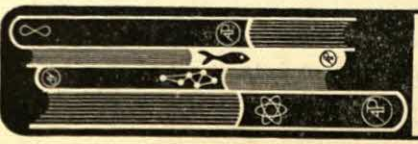
This volume contains contributions by eminent psychologists. Each chapter provides systematic, empirical and theoretical accounts of a specialized topic within the field that has grown out of his own research programs. Much of the data in this book has not been previously published.

CONTENTS:

Abram Amsel, Partial Reinforcement Effects on Vigor and Persistence: Advances in Frustration Theory Derived from a Variety of Within-Subject Experiments
E. J. Capaldi, A Sequential Hypothesis of Instrumental Learning
Harry Fowler, Satiation and Curiosity: Constructs for a Drive and Incentive-Motivational Theory of Exploration
Gordon Bower, A Multicomponent Theory of the Memory Trace
George Mandler, Organization and Memory
AUTHOR INDEX—SUBJECT INDEX.

1967, 381 pp., $11.50

AP 2002



Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

Experiments in Second-Language Learning

By **Edward Crothers**
and **Patrick Suppes**
Both at the Institute for Mathematical Studies in the Social Sciences
Stanford University, Stanford, California

This volume views second-language learning in the light of mathematical learning theory. It describes a series of sixteen experiments involving the acquisition of Russian by native American college students and dealing with phoneme discrimination, phoneme-grapheme correspondence, and the learning of vocabulary, morphology and syntax. The authors formulate mathematical learning models and apply them to the experimental data.

The book is written as a series of technical experimental reports. Discussions of psycholinguistic factors appear separately from the application of models.

CONTENTS:

Some Remarks on Stimulus-Response Theories of Language Learning
Learning to Discriminate Russian Phonemes
Learning the Orthographic Representations of Russian Sounds
Vocabulary Learning
Morphology Learning
Grammar Learning
Suggestions for Future Research
REFERENCES.
AUTHOR INDEX—SUBJECT INDEX.

1967, 374 pp., $13.50

AP 2024

Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

Personal Causation

The Internal Affective Determinants of Behavior

By **Richard de Charms**
Department of Psychology, Washington University
St. Louis, Missouri

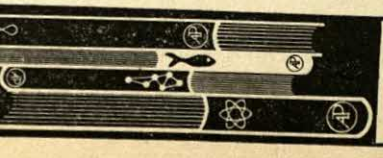
Selecting from the latest published and unpublished research, the author analyzes and integrates the current psychological evidence concerning human motivation within the context of certain philosophical presuppositions of modern behavior theory. The author suggests that the weaknesses of most current theories of motivation result from a paramechanical concept of causation. Viewing man instead as a non-mechanical locus of causality, he discusses affective relationships in animals and man, the measurement of human motives, the social psychological process of attributing motives, and the effects of freedom versus constraint on individual behavior.

CONTENTS:

BASIC PROBLEMS AND PRESUPPOSITIONS
AFFECT, STIMULATION, AND BEHAVIOR
THOUGHT SAMPLING AND ACHIEVEMENT MOTIVATION
PERSONAL CAUSATION
SUMMARY
TEXT
REFERENCES
AUTHOR INDEX—SUBJECT INDEX.

1968, 398 pp., $13.50

AP 2166



Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

INFORMATION FOR AUTHORS

The *Journal of Experimental Child Psychology* will publish papers in which the behavior and development of children is clearly related to its determining variables. Typically (not always) this would mean that a variable has been manipulated in an experimental manner. Articles will be reviewed in terms of the apparent reliability of their findings and the clarity of the relationships they affirm. Research conceived in any theoretical framework is invited, so long as its conclusions are reliable and compelling. Occasionally, critical reviews of related bodies of experimental literature, theoretical discussions which define new paths of experimental work, and well-controlled developmental studies on infra-human subjects will be published. The criterion for acceptance of such papers will be their value in prompting experimental research in related areas with children. Short notes describing new apparatus or techniques, new materials, circuits, or other information valuable to investigators in the field will be accepted from time to time. There is no page charge for publication.

Two copies of each manuscript and each figure or table should be sent to:

Dr. Sidney W. Bijou, Child Behavior Laboratory, 403 E. Healey, University of Illinois, Champaign, Illinois 61820

A third copy should be retained by the author. Manuscripts must be double-spaced and may be typed, dittoed, or mimeographed. Each manuscript must include a concise abstract, preferably not exceeding 300 words, which will appear after the title and author's name in the published article. In some papers this abstract will obviate the need for a summary. In general, authors are free to use whatever organization their paper seems to require, but the specifications of the *Publication Manual* of the American Psychological Association are strongly recommended.

Figures and Tables. Figures should be numbered consecutively with Arabic numerals in order of mention in the text; each figure should have a descriptive legend. Legends should be typed together on a separate sheet, double-spaced. All illustrations should be in finished form ready for the engraver. A convenient size for drawings is $8\frac{1}{2} \times 11$ inches (21×27.5 cm). Drawings should be made with India ink on tracing linen, smooth surface white paper, or bristol board. Graph paper if used should be ruled in blue. Grid lines that are to show in the final engraving should be inked in black. Sharp, clear glossy prints of figures are acceptable.

Tables should be typed on separate pages, numbered consecutively with Arabic numerals in order of mention in the text. All tables should have descriptive headings, typed (double-spaced) above the table.

Literature references in the text should be cited by author's surname and the year of publication, e.g., Smith, 1960; Smith *et al.*, 1959; Smith and Jones, 1961, p. 250 (for references to a specific page). If a reference has more than two authors, the citation includes the surnames of all authors at the first mention, but later citations of the same reference include only the surname of the first author and the abbreviations *"et al."* Suffixes *a, b,* etc., should be used following the date to distinguish two or more works by the same author(s) in the same year, e.g., Smith, 1960a, 1960b. Literature citations in the bibliography should be arranged alphabetically according to the surname of the author.

Journal titles should be written out in full according to the form followed in the Psychological Abstracts (Vol. 39).

The desired style of capitalization and punctuation in references is apparent in the following examples:

STAATS, A. W., AND STAATS, CAROLYN K. A comparison of the development of speech and relating behavior with implications for research. *Child Development,* 1962, 33, 831–846.

SPIKER, C. C. Research methods in children's learning. In P. H. Mussen (Ed.), *Handbook for research methods in child development.* New York: Wiley, 1960. Pp. 374–420.

MUNN, N. L. *The evolution and growth of human behavior.* Cambridge: Riverside Press, 1955.

Reprints: Fifty gratis reprints are supplied and additional copies may be ordered.

A Study of Franz Brentano

His Psychological Standpoint and His Significance in the History of Psychology

By **Antos C. Rancurello**

Department of Psychology
University of Dayton, Dayton, Ohio

CONTENTS:

FRANZ BRENTANO: A PORTRAIT OF HIS LIFE, PERSONALITY AND WORK

OVERVIEW OF BRENTANO'S STANDPOINT IN PSYCHOLOGY

Introduction: Scientific Methodology and the Scope of Human Knowledge. Psychology from the Empirical Standpoint. Genetic or Explanatory Psychology. Descriptive or Phenomenological Psychology.

HISTORICAL PERSPECTIVE

The Emergence and Founding of Psychology (1825-1875): Brentano as an Interpreter of the Zeitgeist.

Early Growth of Psychology (1875-1900): Postscript on Brentano and Wundt and Their Relative Influence upon Psychology during This Period.

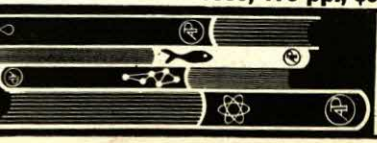
Psychology in Search for Identity (1900-1930): An Overview of Historical Trends and of Their Relationship to Brentano's Standpoint.

Progress of Psychology toward Maturity (1930-Present): Epilogue on Brentano, the "Third Force" Movement, and the Contemporary American Setting in Psychology.

ANNOTATED BIBLIOGRAPHY
AUTHOR INDEX—SUBJECT INDEX.

1968, 178 pp., $8.00

AP 2136



Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

JOURNAL OF Experimental Child Psychology

EDITOR: Sidney W. Bijou

ASSOCIATE EDITORS: Richard S. Bogartz
David Zeaman

EDITORIAL BOARD:

Donald M. Baer
Albert Bandura
Wesley C. Becker
Harry Beilin
Gordon N. Cantor
Peter Eimas
Robert L. Fantz
Eleanor J. Gibson
Eugene S. Gollin
Willard W. Hartup
Raymond H. Hohle
Herbert Kaye
Tracy S. Kendler
Helen L. Koch
Eugene R. Long
David S. Palermo
Herbert L. Pick
Hayne W. Reese
Bryan E. Shepp
Murray Sidman
Einar Siqueland
Patrick Suppes
Joachim Wohlwill

Volume 7, Number 3, June 1969

ACADEMIC PRESS
New York and London

CONTENTS

JOSEPH C. CAMPIONE. Transitivity and Choice Behavior 387

VERNON C. HALL. Acquisition and Transfer Differences between Kindergarteners and Second-Graders on Aurally and Visually Presented Paired-Associates Using an A-B, A-C Design 400

J. ANDRÉ. Reversal Shift Behavior and Verbalization in Two Age Groups of Hearing and Deaf Children 407

LESLIE B. COHEN. Observing Responses, Visual Preferences, and Habituation to Visual Stimuli in Infants 419

L. R. GOULET AND WILLIAM J. HOYER. The Effects of Verbalization on Verbal Discrimination Learning and Associative Recall in Young Children and Adults 434

WILLIAM H. REDD AND J. S. BIRNBRAUER. Adults as Discriminative Stimuli for Different Reinforcement Contingencies with Retarded Children . . 440

GRAHAM M. DAVIES. Recognition Memory for Pictured and Named Objects . 448

JOHN C. YUILLE AND SUZANNE PRITCHARD. Noun Concreteness and Verbal Facilitation as Factors in Imaginal Mediation and PA Learning in Children 459

JOHN S. WATSON AND GARY DANIELSON. An Attempt to Shape Bidimensional Attention in 24-Month-Old Infants 467

L. R. GOULET. The Effect of Verbalization on Discrimination Learning and Transfer in Nursery School Children 479

WALTER L. PORTER. Absolute versus Relational Responding in Children . . 485

RICHARD P. TOISTER AND J. S. BIRNBRAUER. Facilitation of Repeated Free Operant Discrimination Reversals 492

CHARLES Y. NAKAMURA. Effects of Increasing and Decreasing Reward Magnitude and Pre-experimental Persistence Level on Focal and Incidental Responses 514

THOMAS H. SHRINER AND ROBERT L. SPRAGUE. Effects of Time-Compressed Speech Signals on Children's Identification Accuracy and Latency Measures 532

ANNE SARAVO AND EUGENE S. GOLLIN. Oddity Learning and Learning Sets in Children 541

DONALD MEICHENBAUM AND JOSEPH GOODMAN. The Developmental Control of Operant Motor Responding by Verbal Operants 553

THOMAS J. TIGHE AND LOUISE S. TIGHE. Perceptual Variables in the Transposition Behavior of Children 566

STUART M. KEELEY AND ROBERT L. SPRAGUE. Effects of Varying Delay of Reinforcement and Postreinforcement Intervals on Learning of Retardates . 578

AUTHOR INDEX 585

SUBJECT INDEX 587

Published bimonthly at Exchange Place, Hanover, Pa. 17731 by Academic Press, Inc., 111 Fifth Avenue, New York, N. Y., 10003.

In 1969, Volumes 7 and 8 (consisting of 3 issues each) will be published.

Price: $22.00

Private subscriptions (for subscriber's personal use): $10.00

All correspondence and subscription orders should be sent to the office of the Publishers at 111 Fifth Ave., New York, N.Y. 10003

Send notices of change of address to the office of the Publishers at least 4 weeks in advance.

Please include both old and new addresses.

Second-class postage paid at Hanover, Pennsylvania 17331

© 1969 by Academic Press, Inc.

Editorial Note

Since we are publishing two volumes in 1969, the list of psychologists who have served as guest editors for 1969 will appear in the last issue of the second volume, Volume 8. At this juncture we wish to express our gratitude to all those who have cooperated in evaluating the manuscripts.

S. W. Bijou

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 387–399 (1969)

Transitivity and Choice Behavior[1]

JOSEPH C. CAMPIONE[2]
University of Connecticut

A series of three experiments was performed with retarded children in an effort to demonstrate that experimentally induced response strengths are transitive. Experiment I failed to make such a demonstration, leading to the hypothesis that the introduction of a novel transverse pattern into the test series led to the negative results. Experiments II and III, testing with transverse patterns equated for novelty, clearly succeeded in demonstrating transitivity. The close correspondence between data and theory supports the assumption that the underlying scale is a ratio scale. Additionally, Exp. I found that reinforced trials facilitated learning more than did nonreinforced trials.

Learning theories such as those of Hull (1943) and Spence (1956) have assumed that, in a two-choice discrimination, the positive and negative stimuli acquire differential tendencies to evoke a response. The strengths of these tendencies correspond to the reaction potentials of the stimuli. If stimulus A has a greater reaction potential than stimulus B, we should expect *S* to choose stimulus A. As the difference in reaction potentials increases, the likelihood of *S*'s choosing A should increase accordingly.

An assumption implicit in an axiomatic structure of this type is that there exists a scale along which reaction potential can be measured. Further, this scale underlies all of choice behavior, and there should exist laws relating these scale values to their various behavioral correlates. Such a view has been expressed by Hull as follows: ". . . all quantifications of reaction potential, however empirically manifested, should be expressible on the same scale and in terms of the same unit." (Hull, Felsinger, Gladstone, and Yamaguchi, 1947, pp. 238–239.) Certainly, evidence bearing on the existence and properties of such a scale is of no little theoretical significance, for this scale is the very foundation of

[1] Submitted in partial fulfillment of the Ph.D. degree at the University of Connecticut. The author wishes to acknowledge the help of Drs. David Zeaman and Betty House Zeaman throughout the conduct of the research. This research was supported by Grant M-1099 from the National Institute of Mental Health, United States Public Health Service.

[2] Now at the University of Washington.

© 1969 by Academic Press, Inc.

reaction potential (associative strength, response strength) formulations. Further, for such systems to have generality, this scale must number among its properties that of trans-situationality. Specifically, if element A possesses a strength of 10 units in situation one, then it should also possess a strength of 10 units in any other situation in which the boundary conditions of the theory are not violated.

A test of this property of the scale in question requires a derivation which yields a law relating choice situations to one another. One very simple law which follows from the above assumptions is that of transitivity of choice behavior. If an organism prefers (or learns to prefer) stimulus A to stimulus B, and stimulus B to stimulus C, then he should also prefer stimulus A to stimulus C. Further, the strength of this latter preference should be greater than the strength of the stronger of the former two.

The transitivity assumption can be formalized as follows: Let $P(x,y)$ denote the probability that element x will be chosen when x and y are presented together. Then if $P(A,B) \geq .5$, and $P(B,C) \geq .5$, it follows that $P(A,C) \geq max\ [P(A,B),\ P(B,C)]$. Now any theoretical estimate of $P(A,C)$ requires a rule whereby we can relate the strengths of elements to the probability of their being chosen in a given situation. A simple rule is that $P(x,y)$ is a function of the relative strengths of elements x and y.

Luce (1959) has developed a response strength theory of choice behavior in which such a rule is generated. We will be concerned here with two consequences of his Axiom 1. From this axiom, it can be shown that: (a) the v-scale (scale of response strength) is a ratio scale; and (b) $P(x,y) = (v_x) \,/\, (v_x + v_y)$, where v_x and v_y denote the response strengths of elements x and y, respectively. Given this relation, it is simple to derive an expression for $P(A,C)$ in terms of $P(A,B)$ and $P(B,C)$. Letting $p_1 = P(A,B)$, and $p_2 = P(B,C)$, it follows that:

$$P(A,C) = \frac{p_1 p_2}{p_1 p_2 + (1 - p_1)(1 - p_2)} \tag{1}$$

Thus, in addition to testing purely for transitivity (transitivity test), we can also assess the extent to which there is agreement between data and theory on the A–C test (mathematical test). Here close agreement between data and theory would provide inferential support for the assertion that the scale of response strength is a ratio scale. It should also be pointed out that the above equation is not unique to Luce's theory (see Restle, 1961).

The generality of response strength theories is measured by the size and variety of the classes of stimulus elements which are shown to be

well-behaved, i.e., that behave in accordance with the theory. It is strategic to begin a search for well-behaved elements with a demonstration of transitivity, since this property is so basic: nontransitive systems are not fundamentally measureable. Given its importance, it is surprising that this property has not yet been demonstrated for experimentally manipulated response strengths. While a number of studies (e.g., Coombs, 1958) have been concerned with transitivity, they have dealt with preferences rather than experimentally controlled response strengths.

One possible reason for the neglect of such a basic problem may be methodological: transitivity tests require stable *individual* performance curves, while only *group* learning curves emerge from the traditional Trials-by-Subjects design used to study learning. The issue of grouped *vs.* individual data is thus basic to this problem. Several authors (e.g., Hayes, 1953; Estes, 1956; Luce, 1959) have pointed out that learning curves based on group means are not necessarily representative of the functions which would obtain for individual subjects. The traditional Trials-by-Subjects design does not allow such estimation. To handle this problem use will be made of the miniature experiment (a Trials-by-Problems design), derived from the learning set methodology of Harlow (1959) and the miniature experiment of Estes (1960), and adapted for retardate discrimination learning by House and Zeaman (1963).

Experiment I

To test the transitivity assumption, a three-trial design was employed. The first two trials trained Ss to prefer (choose) A over B, and B over C, with the final trial assessing their performance when either A and B, B and C, or A and C were presented. Counterbalancing the order of presentation of the first two trials yields in combination with the three test trials six different experimental problems (EP). A standard problem (SP) was also included in the design which is summarized in Table 1.

TABLE 1
ARRANGEMENT OF STIMULI A, B, AND C FOR SIX EXPERIMENTAL CONDITIONS AND ONE CONTROL

Problem	1	2	3	4	5	6	Standard
Reward value	+ −	+ −	+ −	+ −	+ −	+ −	+ −
Trial 1	A B	A B	A B	B C	B C	B C	A B
Trial 2	B C	B C	B C	A B	A B	A B	A B
Trial 3	A B	B C	A C	A B	B C	A C	A B

Note: The stimuli A, B, and C were selected randomly for each problem for each S. (See text for additional details.)

Note that the SP simply presents the same stimuli on each of the three trials (the standard operations employed in the usual learning experiment). This problem is not used in the test of transitivity, but is included for reasons which will be elaborated below.

Apparatus. A semiautomated version of the Wisconsin General Test Apparatus was used. This apparatus, capable of presenting 12 forms and 7 colors, is described in detail elsewhere (Scott, 1964). The forms were relatively simple geometric figures (e.g., heart, square, triangle). The colors used were red, green, yellow, orange, gray, blue, and violet.

Subjects. Nine institutionalized retardates were selected randomly from the boys and girls attending the Longley School of the Mansfield State Training School. The *S*s had a mean MA of 7–0 years (range: 4–0 to 10–8) and a mean IQ of 56.2 (range: 34 to 82).

Pretraining. *S*s were run to a criterion of 20/25 correct responses on a given day on a standard two-choice form discrimination. The relevant forms were selected randomly for each *S*. Position of reward was determined by a Gellermann (1933) series. A correction procedure was used throughout the experiment. To ensure the fact that *S* was looking at form and form alone, a multivariable irrelevant procedure was used. Briefly, this procedure presented two different colors on each successive trial. Thus, neither color nor color-form compounds could be used as a basis for solution. Since there were only seven colors available, these colors reappeared during any daily session, but with the restriction that no color would be used if it had appeared on either of the two preceding trials.

Procedure. Each *S* was run for 16 daily sessions, a session consisting of three trials on a well-learned SP (the problem which *S* had learned to criterion during the pretraining phase of the experiment), one EP, three trials of the SP, one EP, etc., to a total of six EP. Each daily sessison terminated with three trials of the SP. The forms to be used for each EP were selected randomly subject to the restriction that: (a) the forms used in the SP were never used in an EP; and (b) no forms were used which had appeared on either of the two previous EP. Again the colors to be used on any trial (of either the SP or EP) were selected randomly subject to the restriction that no color would be used if it had appeared on either of the two preceding trials.

During a daily session, each EP appeared once. As nearly as possible, each EP appeared an equal number of times in each serial position throughout the sessions. Also, each of the eight right-left sequences was paired with each EP twice, and with the SP 14 times. During the session, no more than four successive rights or lefts were permitted. Since the design of the experiment caused *S*s to make frequent errors, the

possibility of "failure set" development (Zeaman and House, 1963) presented itself. To minimize this possibility: (a) *Ss* received non-differential reinforcement on Trial 3 (the test trial) of the EP; and (b) the SP (on which *Ss* could perform well) were included. Inclusion of the SP also allowed the assessment of failure set development, high performance indicating a lack of failure set formation.

Results and Discussion

Transitivity. The results are summarized in Table 2, which gives *Ss'* proportions on the three test trials (A-B, B-C, and A-C) separately for each training order. Considering the two training orders separately yields a more rigorous test for transitivity, in that it eliminates averaging error. For example, it would be surprising if *Ss'* probabilities of choosing A over B on Trial 3 were the same following Training Order 1 (Trial 1: A over B; Trial 2: B over C) as following Training Order 2 (Trial 1: B over C; Trial 2: A over B). A recency effect would make the probability of choosing A over B on Trial 3 greater following Training Order 2. In fact, the order of presentation of the two training trials did lead to large differences in perfromance on the test trials (see below).

TABLE 2

PROPORTIONS OF CORRECT RESPONSES ON THE THREE TEST TRIALS OF EXPERIMENT I, AND PREDICTED PROPORTIONS ON THE A-C TEST[a]

	Training order I A B B C					Training order II B C A B				
Subject/ Test	A-B	B-C	A-C	Result	A-C (pred)	A-B	B-C	A-C	Result	A-C (pred)
WL	.25	.69	.56	×	.42	.69	.50	.69	+	.69
JT	.38	.56	.38	×	.44	.69	.44	.44	×	.64
EM	.38	.44	.62	×	.32	.50	.50	.62	+	.50
WF	.50	.56	.56	+	.56	.38	.56	.69	×	.44
GC	.38	.62	.44	×	.50	.56	.56	.62	+	.62
DB	.56	.88	.56	−	.90	.56	.38	.94	×	.44
LK	.44	.69	.69	×	.64	.50	.56	.69	+	.56
MF	.56	.81	.81	−	.84	.44	.69	.56	×	.65
SA	.56	.50	.44	−	.56	.50	.75	.38	−	.75

[a] Proportions are averaged over each training order. Each proportion is based on an *N* of 16.

Examination of Table 2 indicates the absence of the desired transitivity effect, only five of nine possible tests being positive. Note that it is not always possible to test for transitivity, since $P(A,B)$ and

$P(B,C)$ are not both $\geq .5$ (cf. Subject WL, Order 1). A test is counted as positive if $P(A,C) \geq max\ [P(A,B), P(B,C)]$ for $P(A,B)$ and $P(B,C)$ both $> .5$, or if $P(A,C) \geq max\ [P(A,B), P(B,C)]$ for at least one of the pair-wise probabilities equal to .5, with the other $\geq .5$. We can also look at the extent to which the mathematical predictions are borne out. Even in those cases in which we cannot test for transitivity, mathematical predictions can still be made on the A-C test. Again we find little support for the theory, the mean absolute error being 15.3 percentage points. Finally, a chi-square goodness of fit test[3] yielded a χ^2 of 67.46 ($df=18$), with an associated p value, under the null hypothesis, of less than .001. It should be emphasized that the p value is at best approximate, since there is also error involved in estimating $P(A,B)$ and $P(B,C)$.

Reinforcement effects. We can also obtain a measure of the relative strengths of approach and avoidance tendencies by a comparison of performance on the A-B and B-C test trials. When the test trial is A-B, S must make a choice between two alternatives, one of which (A) has been positive once; and one of which (B) has been positive once and negative once; the extent to which A is chosen (or B avoided) is a measure of the strength of avoidance tendencies due to nonreinforcement. In the case of B-C, the choice is between one alternative (B), which has been positive once and negative once; and another (C), which has been negative once. Thus the tendency to prefer B is a measure of the strength of approach tendencies or reinforcement. If $P(B,C) > P(A,B)$, we can conclude that approach tendencies are more powerful than avoidance tendencies. If $P(A,B) > P(B,C)$, the opposite conclusion follows.

As assumption implicit in this argument is that the gain (loss) in response strength due to reinforcement (nonreinforcement) is approximately constant across levels of response strength.

Retention effects. We can also investigate the magnitude of a regency effect in the data. Compare, for example, Problem 1 with Problem 4, and Problem 2 with Problem 5. These differ only in the order of presentation of the first two trials. A recency effect (retroactive interference stronger than proactive interference) would make $P(A,B)$ in Problem 4 greater than $P(A,B)$ in Problem 1 [and $P(B,C)$ in Problem 2 greater than $P(B,C)$ in Problem 5].

[3] The chi-square statistic was calculated using the following formula:

$$\chi^2 = \sum_i \sum_j \left(\frac{[Na_{ij} - N \cdot P_{ij}(A,C)]^2}{N \cdot P_{ij}(A,C)} + \frac{[Nc_{ij} - N \cdot P_{ij}(C,A)]^2}{N \cdot P_{ij}(C,A)} \right)$$

where N = number of A-C test trials (16), Na = number of choices of A, Nc = number of choices of C; i refers to Ss, j to training orders.

The reliability of these effects was assessed by a 2×2 factorial analysis of variance, testing the Recency and Approach-Avoidance variables. The mean percentage of correct responses on the A-B and B-C tests were 50% and 60%, respectively, while the values for the proactive and retroactive interference conditions were 59% and 50%. Both the Recency ($F = 5.19$, $df = 1/32$, $p < .05$) and Approach-Avoidance ($F = 6.49$, $df = 1/32$, $p < .05$) variables attained levels of significance.

The Approach-Avoidance finding is consistent with an earlier experiment (House, Orlando, and Zeaman, 1957) which found approach tendencies to be established more readily than avoidance tendencies for retarded *Ss* with MAs comparable to those of the present study. House and Zeaman (1958) reported the opposite effect for retarded *Ss* with a mean MA of 3½ years. A similar developmental change has been reported with normal children (see Cross and Vaughter, 1966; Evans and Endsley, 1966), performance after nonrewarded trials being superior to that after rewarded trials for children below 4½ years, and the reverse for children above 4½ years.

It is appropriate at this time to mention briefly an earlier version of this experiment. Experiment I was, in fact, a replication of the earlier experiment with one change. In the earlier experiment, the stimuli used were junk pictures (pictures differing multidimensionally in color, form, size, texture). The results of this experiment were quite consistent with Exp. I discussed above. First, transitivity was not demonstrated, only 2 of 10 tests being positive. Second, the same results were obtained with respect to the Recency and Approach-Avoidance variables. It had been hypothesized that the failure to demonstrate transitivity might be due to the use of the highly redundant junk stimuli, with a resulting inability to control *Ss'* attention. Briefly, the argument ran as follows. Assume *S* is limited in attention, and can therefore attend to only a fraction of the component cues on the training trial. If, on the test trial, he now attends to a different subset of the component cues, the effect of the training trial will be lost, and any test of transitivity will be invalidated. This hypothesis, however, finds little comfort in the data of Exp. I.

Another hypothesis also arises. The failure to demonstrate transitivity in these two experiments may be due to the use of both novel and familiar transverse patterns (Spence, 1952) in the test situation. A transverse pattern is that pattern formed by the two stimuli presented on a trial. Thus, when A is trained over B, the transverse pattern formed by A and B is also present, and some learning can proceed with respect to this transverse pattern itself. For example, *S* may learn to "approach A, when A and B are presented together."

In Exp. I, three types of test trials were used. Two of these, pairs

A and B, and B and C, were simply replicas of either Traning Trial 1 or Training Trial 2. However, the third pair, consisting of A and C, were never presented together in training so that the transverse pattern defined by A and C was a novel one. It seemed possible, then, that the novel transverse pattern might have affected performance differently from that of the familiar transverse patterns, confounding the comparison made among the three types of test trials. Note that if this hypothesis is correct, however, an additional complexity presents itself. The differential effect (if any) of novel and familiar transverse patterns implies that the response strengths are to an extent not trans-situational. Controlling for novelty then becomes a theoretical boundary condition. Difficulties in demonstrating transitivity, then, while not a disproof of the underlying theoretical structure, provide a measure of the generality of the theory. This generality is assessed by the size and variety of classes of stimuli which demonstrate the properties ascribed to them by the theory. Any boundary condition, such as the one above, would clearly restrict this generality.

There are two ways in which the novelty of the test patterns can be equated: (a) all of the test trials may be new transverse patterns; and (b) none of the test trials may be new transverse patterns. Experiments II (using the former procedure) and III (the latter), employing the stimuli and procedure of Exp. I, were designed to control this source of variance.

Experiment II

To control for the confounding novelty effect, a different training sequence was used. None of the stimuli used in the ensuing tests (A, B, and C) were ever paired on a training trial. To accomplish this a fourth stimulus (D) was introduced. The training sequence then consisted of the following five trials: three trials on which A was trained over D, one trial on which B was trained over D, and one trial on which D was trained over C. Thus A was reinforced three times, B was reinforced once, and C was extinguished once. The training sequence was followed by the three test trials (A-B, B-C, and A-C), all of which were novel transverse patterns.

Subjects. Ten institutionalized retardates were selected randomly from the boys and girls attending the Longley School of the Mansfield State Training School. They had a mean MA of 8–11 years (range: 7–8 to 10–6) and a mean IQ of 61.7 (range: 48–70).

Procedure. The apparatus and pretraining were identical to those of Exp. I. There are 5!/3! 1! 1!, or 20, possible orders of the five train-

ing trials and six orders of the three test trials, yielding a total of 120 eight-trial EP. These were divided into two sets of 60 EP, each set consisting of ten of the training orders in combination with the six orderings of the test trials. A given *S* received one set of these problems. In this way, each test pair probability per *S* was based on 60 observations. Four of the eight position sequences were paired with the test trials eight times and the other four seven times. Position of reward for the training trials and SP (see below) was random subject to the following restrictions: (a) the number of repetitions (reinforcement of the same side on two successive trials) in going from the last training trial to the first test trial was equal to the number of nonrepetitions.

Each daily session consisted of one SP (five trials), one EP (eight trials), one SP, one EP, one SP, one EP, and one SP, for a total of seven problems (44 trials) per day. Two sets (one from each set of 60 EP) of 20 sessions were constructed, and *S*s were randomly assigned one set or the other. Each *S* was then given the 20 sessions assigned him in a random order. During each session, the forms used on the first SP were selected randomly for each *S*. The forms for the six succeeding problems were chosen randomly subject to the restriction that they had not appeared on any of the three preceding problems. The multivariable irrelevant procedure was used throughout, again color being the irrelevant dimension. A correction procedure was used throughout the experiment.

Results and Discussion

The results are summarized in Table 3. Here the transitivity effect is very strong, occurring with each *S*. Moreover, the correspondence

TABLE 3

PROPORTIONS OF CORRECT RESPONSES ON THE THREE TEST TRIALS OF EXPERIMENT II, AND PREDICTED PROPORTIONS ON THE A-C TEST*

Subject	A-B	B-C	A-C	A-C (pred)
JD	.60	.57	.68	.67
MB	.57	.62	.68	.69
RC	.58	.55	.62	.63
CM	.62	.58	.72	.69
CG	.60	.63	.75	.72
SR	.67	.68	.90	.81
FW	.77	.68	.88	.88
LL	.62	.65	.78	.73
PM	.70	.72	.88	.86
PL	.57	.63	.75	.69

* All proportions are based on an *N* of 60.

between data and the mathematical prediction is striking (predictions were obtained from Equation 1).

A test of the mathematical predictions was carried out via a chi-square goodness of fit test, which yielded a χ^2 of 4.918 ($df = 10$), with an associated p value, under the null hypothesis, of greater than .90. Again, the p value is at best approximate. Finally, a systematic error is present in the data. For the seven cases in which there was a measureable error, *S* always performed at a higher level than was predicted. The probability of all seven errors being in the same direction is less than .02.

The increased strength of the transitivity effect over Exp. I is attributed to the use, on the test trials, of transverse patterns equated for novelty. These results are quite significant for response strength theory, since they furnish evidence for the assumption that response strengths do form a well-behaved scale and that, within certain boundary conditions, these scale values are trans-situational. The close agreement between data and the mathematical predictions further lends indirect support to the assumption that there exists a ratio scale along which response strengths can be measured.

If the poorer results of Exp. I are due, as now seems likely, to failure to equate the novelty of the transverse patterns used on the test trials, then Exp. III should be expected to yield positive results once again.

Experiment III

The objective of Exp. III was to eliminate confounding due to transverse pattern novelty by using three familiar patterns on the test trials. This was accomplished by using three training trials: A over B, B over C, and A over C, followed by a test on one of the three relations.

Subjects. Eight institutionalized retardates from the same population served as *Ss*. A scarcity of *Ss* in the higher MA range led to the inclusion, in Exp. III, of four of the *Ss* who had served in Exp. II. The eight *Ss* had a mean MA of 9-4 years (range: 7-9 to 10-11) and a mean IQ of 68.6 (range: 56 to 80).

Procedure. The apparatus and pretraining were identical to those of Exps. I and II. There are six possible training sequences (of A-B, B-C, and A-C) and three possible tests, yielding a total of 18 4-trial EP. Each problem was used five times, making a total of 90 problems per *S*. Thus, each *S* was given 30 tests on each of the pairwise comparisons. A daily session consisted of one SP (four trials), two EP, one SP, two EP, one SP, two EP, and one SP for a total of 10 problems (40 trials) per day. On each day, the six EP contained two tests on each of the three relations. Each of the test pairs appeared an equal number of

times (five) in each of the six positions throughout the daily sessions. For the 4-trial problems, there are 16 possible sequences of right and left. Eliminating those in which all four are to the same side (a total of two) and those in which three successive are to the same side (four) leaves 10 sequences. The pairing of position sequences and EP (specifically, test trials) was counterbalanced. Further, the number of repetitions from Trial 3 to Trial 4 (the test trial) equalled the number of nonrepetitions for each of the test pairs. These 10 position sequences were applied to the SP six times each. The final restriction on the position sequences was that, on any given day, there could not be more than four successive trials on which the same side was rewarded. A total of 15 daily sessions was prepared, and each *S* went through them in a random order. As in Exps. I and II, the multivariable irrelevant procedure was used throughout. The forms used on the first SP of a daily sequence were selected randomly for each *S*. The forms on the remaining nine problems were also selected randomly subject to the restriction that they had not appeared on any of the three preceding problems. Colors were selected in the same way as during Exps. I and II.

Results and Discussion

The results are summarized in Table 4. Again the transitivity effect is strong, being shown by all *Ss*. Close correspondence also exists between the data and mathematical predictions. A chi-square goodness of fit test yielded a χ^2 value of 6.126 ($df = 8$), with an associated p value of greater than .60. Further, for only one *S*, CW, was there any significant difference between the observed and predicted values. It should also be noted that no systematic error appears in the data, four of the

TABLE 4

PROPORTIONS OF CORRECT RESPONSES ON THE THREE TEST TRIALS OF EXPERIMENT III, AND PREDICTED PROPORTIONS ON THE A-C TEST*

Subject	A-B	B-C	A-C	A-C (pred)
FW	80	77	87	88
GC	76	60	77	73
GCv	70	70	80	84
RR	60	73	77	90
LL	57	67	70	73
CW	50	57	77	67
CG	63	60	73	73
RR	50	70	70	70

* All proportions are based on *N* of 30.

measurable errors being positive and three negative. A chi-square goodness of fit test, combining Exps. II and III, yielded a χ^2 value of 11.044 ($df = 18$), and a p value greater than .85. Again, these p values can be considered only approximate.

These results are consistent with those of Exp. II and strengthen the argument that the failure to demonstrate transitivity in Exp. I was due to the fact that the transverse patterns used on the three types of test trials were not equated for novelty.

With respect to transitivity, these experiments contribute three major findings: (a) transitivity of response strength is demonstrated; (b) evidence of a ratio scale of response strength is inferred; and (c) a boundary is indicated on the conditions under which transitivity of response strength can be expected to hold.

REFERENCES

Coombs, C. H. On the use of inconsistency of preferences in psychological measurement. *J. of Experimental Psychology*, 1958, **55**, 1–7.

Cross, H. A., and Vaughter, R. M. The Moss-Harlow effect in preschool children as a function of age. *J. of Experimental Child Psychology*, 1966, **4**, 280–284.

Estes, W. K. The problem of inference from curves based on grouped data. *Psychological Bulletin*, 1956, **53**, 134–140.

Estes, W. K. Learning theory and the new "mental chemistry." *Psychological Review*, 1960, **67**, 207–223.

Evans, P. K., and Endsley, R. C. Effects of three reward-nonreward ratios on discrimination learning in preschool children. *J. of Experimental Child Psychology*, 1966, **4**, 143–149.

Gellermann, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. of Genetic Psychology*, 1933, **42**, 206–207.

Hayes, K. J. The backward curve: a method for the study of learning. *Psychological Review*, 1953, **60**, 269–275.

Harlow, H. F. Learning set and error factor theory. In S. Koch (Ed.) *Psychology: A study of a science*, Vol. 2. New York: McGraw-Hill, 1959. Pp. 492–537.

House, B. J., and Zeaman, D. Reward and nonreward in the discrimination learning of imbeciles. *J. of Comparative and Physiological Psychology*, 1958, **51**, 614–618.

House, B. J., and Zeaman, D. Miniature experiments in the discrimination learning of retardates. In L. Lipsitt and C. C. Spiker (Eds.) *Advances in child development and behavior*, Vol. 1. New York: Academic Press, 1963. Pp. 313–374.

House, B. J., Orlando, R., and Zeaman, D. Role of positive and negative cues in the discrimination learning of mental defectives. *Perceptual and Motor Skills*, 1957, **7**, 73–79.

Hull, C. L. *Principles of behavior*. New York: Appleton, Century-Crofts, 1943.

Hull, C. L., Felsinger, J. M., Gladstone, A. I., and Yamaguchi, H. F. A proposed quantification of habit strength. *Psychological Review*, 1947, **54**, 237–254.

Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.

Restle, F. *Psychology of judgment and choice*. New York: Wiley, 1961.

Scott, K. An apparatus for the automatic presentation of color and form. *J. of Experimental Analysis of Behavior*, 1964, **7**, 205–206.

SPENCE, K. W. The nature of the response in discrimination learning. *Psychological Review*, 1952, **59**, 89–93.

SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale University Press, 1956.

ZEAMAN, D., AND HOUSE, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.) *Handbook of mental deficiency*. New York: McGraw-Hill, 1963. Pp. 159–223.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 400–406 (1969)

Acquisition and Transfer Differences between Kindergarteners and Second-Graders on Aurally and Visually Presented Paired-Associates Using an A-B, A-C Design

VERNON C. HALL[1]
Syracuse University

Kindergarten and second-grade children were trained in an A-B, A-C paired-associate task using five pairs and two modes (aural and visual) of presentation. It was found that, contrary to expectations, acquisition on the first list was significantly faster for the visual mode. In addition, both Kindergarteners and Second-graders exhibited significant negative transfer.

As Jensen and Rohwer (1965) point out, many conceptions about the learning process might be considerably different if psychologists had used *Ss* other than college sophomores in rote learning experiments. It is interesting that while many psychologists hypothesize developmental changes (be they maturationally or experientially based) in other areas, rote learning has until recently been both theoretically and empirically neglected in this regard. The present study was carried out to extend our knowledge about developmental differences in paired-associate learning. More specifically, this *E* was concerned with the effects of mode of presentation on acquisition and interference with kindergarten and second-grade children.

In an earlier article, Loomis and Hall (1968), using the aural mode of presentation, employed the A-B, A-C design with kindergarten and second-grade children. They found significant interference effects with the 8-year-olds but not the 5-year-olds. This is not only contrary to White's (1965, p. 214) prediction that 5-year-olds should exhibit significantly more negative interference, but also fails to demonstrate a well-known phenomenon found many times with adults. This clearly demanded replication with a new sample. In addition, most studies utilizing the A-B, A-C paradigm have used visual presentation and it was felt that adding the visual mode here would considerably increase our

[1]Present address: Psychology Department, Syracuse University, Syracuse, New York 13210.

knowledge about the generality of the Loomis and Hall findings. The Loomis and Hall study also used a double criterion (15 trials or 1 perfect trial) for its measurement of transfer effects and running everyone the same number of trials should again increase our confidence in the results. Finally, an additional pair was added to the lists (from four to five), since it was found that the younger children were easily able to handle this many.

With regard to acquisition, it has typically been reported that young children perform better with aural than visual presentation (e.g., McGeoch and Irion, 1952). Budoff and Quinlan (1964a), using 8 pairs of nouns and verbs from preprimers, some of which formed meaningful combinations (e.g., look–dog) found that 7- and 8-year-olds learned the lists significantly faster when these were presented aurally. They also used a double criterion of 1 perfect trial or 24 trials. Later, the same *E*s (1964b) replicated these findings using average and retarded readers with the same materials and procedures. Otto (1961) paired five common forms (e.g., triangle) with low association value trigrams which he presented to three grade levels (2, 4, and 6) including good, average and poor readers. Here the difference between procedures used in aural and visual presentation (he calls them reinforcements) modes are less pronounced (in both conditions the cvc was articulated, but the visual condition included presentation of the three letters). Here, again, the second-graders (but not fourth or sixth) were better at the aural than visual presentation. There was an added problem, however, in that both serial order and scrambled order were used and *S* was run until he reached one perfect trial for each presentation type.

In no case have *E*s used kindergarten children or pictures rather than words in comparing visual with aural presentations. Since, as White (1965) has pointed out, the age range between 5 and 7 has been identified by several theorists and much empirical data as being a transitional stage for learning processes, it is especially appropriate for any *E* interested in early developmental changes to include this age span. With regard to the use of pictures rather than words, not only is it impossible to use the latter with kindergarten children, but it seems more appropriate to use pictures with second-graders, since ability to read might well be a confounding factor when making a comparison with the aural presentation.

METHOD

Design and sample. The basic design compared two grade levels (kindergarten and second grade), two modes of presentation (visual and aural), and two paradigms (experimental and control). The control

group learned the standard D-E, A-C list. The *Ss* were 60 kindergarteners randomly chosen from about 125 students and 60 second graders randomly chosen from 130 students at Wetzel Road Elementary School, Liverpool, New York. The mean ages were 6.1 and 8.2 years, respectively. The *Ss* in each age group were randomly assigned to one of the four groups (aural or visual; experimental or control), resulting in the inclusion of 15 *Ss* in each cell.

Material and procedures. The lists used are shown in Table 1 and were made up of nouns randomly selected from the 45 singular nouns used as stimuli in gathering the oral word association norms from young children by Palermo and Jenkins (1966). Outline drawings of the objects these words represented were then made, photographed, and reproduced as 35mm slides. These slides were projected on a 9 × 9-inch screen by a Sawyer 707Q slide projector. The projector was operated automatically by two Hunter Timers. The first object was presented for 3 seconds by itself and then the two objects were shown together for 3 seconds. The intertrial interval was 6 seconds. The aural presentation was recorded on a Magnacord 1140 Tape recorder and presented to the children via Sharpe earphones. The stimulus word was first pronounced and then after a 3-second pause it was pronounced again with the response word. Both modes were in all other details identical. The actual lists were arranged in five random orders with the limitation that no single pair was to occur consecutively. This was violated on Trial 10 where the circular tray on the slide projector recycled and the last pair on Trial 9 was the same as the first pair on Trial 10.

Prior to the experimental task, each *S* was presented a two-pair list (knife-fork and ice cream-cake) to a criterion of one perfect trial to be sure the children understood the instructions. The instructions seen below were adapted from McCullers (1963).

> This is a game to see how well you can learn. If you try hard, we will give you some M&M candy afterwards. This is how it goes. When I turn on this slide projector (tape recorder), you will see (hear a word) an object. A few seconds later, you will see (hear) the same object (the same word) with another object (word). Your job is to guess what the second object (word) will be, before you see (hear) it. You are to say the name of the object (word) out loud so I can hear you. *Although the objects are all very common, I will tell you what we call them the first time through if we happen to call them different things.* Here is an example of what you are going to see (hear). (1 trial with the two pair list) O.K. now we are ready to start. (Present the two pair list until *S* gets one perfect trial and then clarify instructions if necessary.)

The italicized sentence was only used in the visual condition. Other condition differences are in parentheses and should be self-evident. The

TABLE 1
WORD LIST USED

DE	AB	AC
Doctor-chair	House-gun	House-bread
Table-hammer	Finger-woman	Finger-sheep
Bed-dog	Needle-cheese	Needle-girl
Spider-man	Lion-scissors	Lion-boy
Hand-lamp	Baby-car	Baby-mountain

candy was used to keep the attention of the children and each *S* received four M & M candies upon completion of the task.

All *S*s repeated the first list until one perfect anticipation trial was performed and then immediately shifted to the second (within 30–60 seconds). The second list was presented to all *S*s for 9 trials. Table 1 shows the lists used.

RESULTS

Tables 2 and 3 which show the appropriate acquisition information for the various lists indicate that the correlations between the means and standard deviations are quite high. A subsequent examination of the distributions indicated that they were skewed. Although it is now generally accepted that the analysis of variance is relatively insensitive to this problem (see Box, 1953) all data analyses were done using both raw scores and log transformations. As would be expected, while the correlations were reduced considerably, the results remained unchanged. *F* values given here were those obtained using raw scores.

A 2 (experimental-control) $\times$ 2 (kindergarten-second grade) $\times$ 2 (audio-visual) analysis of variance was computed on trials to criterion on the first list. This analysis yielded a significant visual-audio effect ($F = 5.11$, 1/112 *df*, $p < .05$) and a significant experimental-control interaction with kindergarten and second grade ($F = 4.50$, 1/112 *df*,

TABLE 2
MEANS AND STANDARD DEVIATIONS FOR NUMBER OF TRIALS TO CRITERION ON THE FIRST LIST

	Experimental		Control	
Group	Mean	*SD*	Mean	*SD*
Kindergarten audio	7.8	3.95	11.2	5.52
Visual	7.2	4.57	7.1	2.94
Second grade audio	8.1	4.67	6.7	3.49
Visual	7.3	3.45	5.7	2.49

$p < .05$). Table 2 indicates that children reached criterion on the list faster when it was visually presented. The interaction was caused by the kindergarten controls taking longer than the experimental group while the second grade controls did better than the experimental group. Separate t tests for each age were computed comparing the experimental and control groups. Both comparisons failed to reach significance. No other interactions or main effects were significant.

The second list was analyzed with a 2 (experimental-control) $\times$ 2 (aural-visual) $\times$ 2 (kindergarten-second grade) analysis of variance using number of correct anticipations. The means and standard deviations are shown in Table 3. This analysis yielded a significant experimental-control effect ($F = 11.11$, $p < .01$, 1/112 df) and a significant presentation mode by age interaction ($F = 4.14$, 1/112 df, $p < .05$). This was the result of age differences in performance on the visual mode of presentation. The visual mode seemed to be easier for the second graders. No other main effects or interactions were significant.

TABLE 3
MEANS AND STANDARD DEVIATIONS OF NUMBER CORRECT ON TRIALS 2–9 ON SECOND LIST

	Experimental		Control	
Group	Mean	*SD*	Mean	*SD*
Kindergarten audio	23.33	8.24	27.33	6.84
Visual	20.87	10.50	25.33	8.95
Second grade audio	23.00	10.31	26.73	7.59
Visual	24.67	8.71	33.33	6.99

DISCUSSION

Although it is, of course, impossible to exactly equate the two modes of presentation, the writer feels that by using pictures rather than printed words and earphones rather than free field, he has come closer than previous attempts. The present writer also feels that in light of these findings the earlier generalization (McGoech and Irion, 1952; Budoff and Quinlan, 1964a) that children learn faster aurally than visually must be modified. Most studies upon which these generalizations were based used older *S*s than the present study. It seems possible that the conditions under which different modes excel may be quite specific, with both *S* and task parameters in need of further investigation.

It again appears that the prediction made by White (1965) has failed to materialize. The fact that Loomis and Hall (1968) failed to find a significant interference effect with slightly younger children using the

aural presentation mode (mean of 6.1 as opposed to 5.9) indicates that the findings tend to be opposite from the predicted direction.

The fact that the Loomis and Hall findings did not replicate could be explained in several ways and need further investigation. First, the addition of a non-contingent candy reward could have increased the attending behavior of the younger subjects who may have not yet been used to paying attention for long periods of time. The candy may also account for the nonsignificant age main effects which were found in the earlier study. In addition, earphones were used in the present study, which might have also influenced attention in the aural mode.

Perhaps the most pressing problem concerns the implications such findings have for how these processes operate in young children. This may be phrased in terms of questions asked by stage theorists or more simply by people involved in mediation studies. While there have been a number of findings which suggest that 5-year-olds perform differently from older children (i.e., Koppenaal, Krull, and Katz, 1964; Jensen and Rohwer, 1965) the tendency so far has been to account for them simply in terms of differences in experience with language and words. Future studies need to systematically investigate more specifically the kinds of experiences which are hypothesized to be important before this question can be fully settled.

ACKNOWLEDGMENT

The author thanks Arlene Blake, Richard Kingsley, and Gwenn Simpson for their valuable assistance. This research was supported by Contract OEO-4120 U. S. Office of Economic Opportunity and by Contract II OEC-1-7-0700632834, U. S. Office of Education.

REFERENCES

Box, G. E. P. Nonnormality and tests on Variance. *Biometrika*, 1953, **40**, 318–335.

Budoff, M., and Quinlan, S. Auditory and visual learning in primary grade children. *Child Development*, 1964a, **35**, 583–586.

Budoff, M., and Quinlan, D. Reading progress as related to efficiency of visual and aural learning in the primary grades. *Journal of Educational Psychology*, 1964b, **55**, 247–252.

Jensen, A. R., and Rohwer, W. D. Syntactical mediation of serial and paired-associate learning as a function of age. *Child Development*, 1965, **36**, 601–608.

Koppenaal, R. J., Krull, A., and Katz, H. Age, interference, and forgetting. *Journal of Experimental Child Psychology*, 1964, **1**, 360–375.

Loomis, D., and Hall, V. The effects of age and overlearning in an A-B, A-C design using high and low H value. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 182–188.

McCullers, J. C. An analysis of some factors underlying intralist associative transfer in paired-associate learning. *Journal of Experimental Psychology*, 1963, **65**, 163–168.

McGeoch, J. A., and Irion, A. L. *The psychology of human learning.* New York: David McKay Company, Inc., 1952.

Otto, W. The acquisition and retention of paired associates by good, average, and poor readers. *Journal of Educational Psychology,* 1961, **52,** 241–248.

Palermo, D. S., and Jenkins, J. J. *Oral word association norms for children in grades one through four.* Research Bulletin No. 60, Dept. of Psychology, The Pennsylvania State University, University Park, Pennsylvania, 1966.

White, S. H. Evidence for a hierarchial arrangment of learning processes. In L. P. Lipsitt and C. C. Spiker (Eds.) *Advances in child development and behavior.* New York: Academic Press, 1965.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 407–418 (1969)

Reversal Shift Behavior and Verbalization in Two Age Groups of Hearing and Deaf Children

J. ANDRÉ[1]

University of California, Berkeley

Four groups each containing 24 children (hearing or deaf, 7 or 12 years of age) were given the Kendler "optional choice" reversal-nonreversal shift task to determine the relationship between language ability and performance in this task. Following completion of the task, *Ss* were asked to verbalize the concept.

The results indicate no difference between hearing children and their deaf counterparts in either speed of learning the first or second concept, or in the proportion of "reversers," but there was a significant difference between the two ages for all these measures.

For the hearing children, verbal reports were not significantly related to task performance. Further analysis demonstrated a significant relationship between verbal report and speed of shifting to the second concept as well as between type of shift and speed of shifting.

Numerous recent studies have been concerned with the theoretical interpretation of the results of reversal-nonreversal shift tasks given to human *Ss*. In these studies it has been found that preschool children learn a nonreversal faster than a reversal shift task (Kendler, Kendler and Wells, 1960), whereas college students learn a reversal faster than a nonreversal shift (Kendler and D'Amato, 1955). In studies in which each *S* was allowed to make either shift, they found that the percentage of *Ss* choosing to make a reversal shift increased with age, from 37.5 at 3 to 62.5 at 10 years (Kendler, Kendler and Learnard, 1962). Reasoning that a reversal should be faster than a nonreversal shift only if the *Ss* were able to use some mediator, they have defined the presence of a mediator by the *S*'s choice on the optional shift task. *Ss* who choose to make a reversal (R) shift are said to be exhibiting mediated behavior, and are

[1] Now at Sacramento State College. An early version of this paper was presented at APA 1964. Appreciation is expressed to the following schools for the generous cooperation afforded to the *E* in providing subjects and space: California School for the Deaf, Mr. M. A. Leenhouts, Assistant Superintendent; Hawthorne School, Oakland, Mr. L. Brennan, Principal; Whittier University School, Berkeley, Dr. R. D. Muscio, Principal.

termed "mediators," whereas those who make a nonreversal (NR) shift or respond inconsistently (In) are termed "nonmediators."

The Kendlers have suggested that the most obvious and reasonable source of this mediation is verbal behavior (H. H. Kendler, 1954, p. 350; 1961, p. 203; T. S. Kendler, 1963, p. 40, 42, 44; 1964, p. 428) and have presented considerable evidence supporting this position. Kendler *et al.* (1960) found that giving one group of nursery school children explicit instructions to verbalize while performing the task had no effect on their performance in a forced choice task. In another study (Kendler and Kendler, 1961) 4- and 7-year olds were given either relevant, irrelevant or no verbalization instructions prior to a R shift task. It was found that verbalization instructions significantly influenced the performance for both age groups combined, the greatest effect being the irrelevant verbalization condition for the older children.

T. S. Kendler (1964) gave the optional choice R shift task to kindergarten children in two separate experiments to determine if verbalization by the children while they were learning would increase the frequency of those choosing to make a R shift. In both experiments, verbalization significantly increased the frequency of R choice behavior, though she notes that in neither experiment was there any evidence that verbalization affected the speed of learning the optional shift. It is possible that requiring the children to respond to both stimuli on each trial (*Black* is the winner, *White* is the loser) built up the association between Black and White as relevant attributes. When Black is no longer correct in the second stage, White is chosen because of the associative strength to Black or because verbalizing Black and White increases attention to the brightness dimension. Either possibility would facilitate the R shift.

While these studies indicate that verbalization as an independent variable can have an effect on the results of both forced and optional choice R shift tasks, two questions still remain: (1) Is mediation in this paradigm *necessarily* verbal? and (2) Is mediation related to verbal ability when verbalization is not introduced as an independent variable?

If we reason that this mediation is due to verbal mechanisms, a possible test would consist of giving the "optional choice" shift task to two groups who differed in their verbal abilities but were equated for age and I.Q. Since deaf children lag considerably behind their hearing counterparts in verbal abilities of almost every conceivable type (Ewing and Ewing, 1938; Lenneberg, 1967), it was decided that they would provide a reasonable source of evidence relating to this theory. Therefore, two age groups (7 and 12) of hearing and deaf children were selected as subjects in this experiment.

In order to determine the relationship of verbal ability to task per-

formance for both deaf and hearing *Ss*, all *Ss* were given a posttask quiz to assess their ability to describe the correct stimuli to which they had been responding.

METHOD

Subjects

The deaf *Ss* were 24 children in each age group (a 12-year age group with a mean age of 142 months and a range from 138 to 147 months, and a 7-year age group with a mean of 83 months and a range from 71 to 91 months) from the California School for the Deaf in Berkeley, and Hawthorne School in Oakland. The 12-year group had a mean IQ of 101.4 with a range from 83 to 128 (based on available data for 18 children and a variety of tests including the WISC, Columbia, Good-enough, and Leiter). There were no measures for the 7-year group, but since all available children were used in the study and the entrance examination to these institutions excludes those of gross subnormal intelligence, it is quite likely that these are a fairly normal group. The 12-year group had 10 girls and the 7-year group had 12 girls.

Hearing *Ss* were 24 normal hearing children in each of two age groups (a 12-year group with a mean age of 143 months and a range from 134 to 151 months, and a 7-year group with a mean age of 82 months and a range from 75 to 88 months) from the Whittier University School in Berkeley. The 12-year group had a mean IQ of 102.2 with a range from 77 to 117, based on Henman-Nelson tests given the previous school year. The members of the 7-year group were chosen by their teachers as average normal pupils being neither very bright nor very dull. The 12-year group had 9 girls and the 7-year group had 15 girls.

Ten *Ss* not included above were eliminated. Each 12-year group lost one because of experimenter error or interruption, and each 7-year group lost two for the same reason. One 7-year hearing and three 7-year deaf children were eliminated when they could not learn the original discrimination within 280 trials.

Apparatus

The discrimination apparatus was as similar as possible to that used by Kendler *et al.* (1962), consisting of a box $17 \times 10 \times 4$ inches with paired windows, levers and reward openings arranged symmetrically on the front. Windows were 3×3-inch openings behind which the stimulus cards were placed. Choice of one stimulus or another was made by moving the proper brass rod up and toward the stimulus. The correct choice was accompanied by a loud click of the electric marble-magazine

resulting in the delivery of a marble into the reward opening. Stimuli were four medium-gray cardboard cards mounted with squares of different size and brightness: One large (3 square inches) black (B), one small (1 square inch) black (b), one large white (W), and one small white (w). These were paired so that at each presentation two cards appeared which varied simultaneously on two dimensions. It should be noted that there were only two possible combinations which vary thus: B paired with w, and W paired with b. A sliding screen was lowered in front of the windows while the stimuli were being changed by *E*.

Procedure

Ss within each group were assigned random numbers to determine the sequence in which they were tested. Both ages and conditions were run concurrently. All possible stimulus combinations were counterbalanced for each group.

Ss were run individually in a room with only the *E* present. Deaf *Ss* had the task explained to them by a skilled teacher at first and later by *E* when he became proficient and no teacher was available. The instructions given the *Ss* were as close as possible to those used by Kendler, *et al.* (1962). The child was told that he was going to play a game with *E* in order to win marbles for correct choices. Hearing *Ss* were given many of the same gestures used in the explanation for the deaf. All children were told to "look and think" and they would be able to win many marbles. They were given occasional encouragement throughout the first task to keep up their interest.

All but four *Ss* completed the experiment in one session. Two 7-year deaf children, one 12- and one 7-year hearing children required two sessions.

Three series of trials were run with no break between them. Series I gave training in the original discrimination, Series II gave training in a second discrimination and Series III consisted of five test trials interspersed with five series II trials.

In Series I, both pairs (B vs. w and W vs. b) were presented alternately in a fixed irregular pattern so that each card appeared an equal number of times on each side with no more than two in succession except where absolutely necessary to break a position habit after 60 trials had been given. The reward pattern was such that one card of each pair was always correct, and the same concept (large, small, black or white) was always correct for any given *S*. A marble was automatically delivered for a correct choice, but for an error, *S* was required to return a marble to *E*. Five marbles were given to *S* at the beginning. If *S* lost all the marbles, he was told that he "lost" the game and was given five more to

play a "new" game. *S*s were run to a criterion of 9 out of 10 consecutive correct responses. Any *S* requiring more than 280 trials to learn this initial discrimination was rejected and replaced by an equivalent *S*.

In Series II only one pair was presented repeatedly (B vs. w *or* W vs. b) and the reward pattern was reversed from Series I; i.e., if large had been rewarded in Series I, small was rewarded in Series II. Since there was no second pair to indicate which of the two dimensions was relevant, the *S* could respond to either or both. For example, suppose that in Series I large was rewarded (W+ vs. b— and B+ vs. w—). Now, in Series II small is rewarded, but since only one pair of stimuli is being presented; i.e., B— vs. w+, *S* could reach criterion by responding to small, white or both. Series III was given to determine the dimension to which *S* was responding.

In Series III, both pairs were presented alternately, each five times, the pair omitted during Series II being the test pair. Both stimulus cards not appearing in Series II were rewarded while the pair presented in Series II retained the same reward pattern. The first pair presented in Series III was always a test pair and was always presented as it had been on the first trial of Series I. Thus, all possible cards and positions were counterbalanced for the test pair across *S*s. Going back to the above example: In Series I, large was rewarded and in Series II small was rewarded (but white was also). The reward pattern for Series III would then be: W+ vs. b+ and B— vs. w+. Therefore, no matter which dimension *S* was responding to in Series II, the same dimension would be rewarded in Series III. This series was given to determine the dimension to which the response was being made, and thus identify reversers and nonreversers. In the Kendler *et al.* (1962) study, 10 test trials were given, and a reverser was defined as an *S* who chose the reversal choice at least 8 out of 10 *test* trials. Since only 5 *test* trials were given in this study, the corresponding definition would be 4 out of 5. Therefore, if the *S* chose small in at least 4 of the 5 test trials he was classified as a reverser. If he chose white in at least 4 of the 5 test trials, he was classified as a nonreverser, and if he chose any single concept less than four times he was classified as inconsistent.[2]

[2] The use of 4/5 instead of 8/10 as the basis for classification of the type of response made in Series III might appear to present a problem in increased misclassifications. However, most *S*s respond in a consistent fashion in Series III and this appears to be only a minor problem. In a recent study ten test trials were presented for each of the 96 seven-yr. hearing children. Reanalyzing these data on the basis of the first five compared to the entire ten test trials revealed nine misclassifications which is approximately what would be expected with the 8/10 criterion if *S*s were operating on a random basis. Hence, *S* are generally consistent and the 4/5 criterion does not create a major problem of misclassification.

After the test series was completed and the stimulus cards covered each *S* was asked "which picture is the right one?" If the *S* did not answer verbally, the test pair was shown to him and the question was repeated. *Ss* often would point at the correct card, but *E* urged them to verbalize the correct answer. Only verbal answers were accepted and in the case where several dimensions were verbalized, the first one mentioned was always accepted as the answer. These responses were classified as correct if they matched their performance in Series III. Verbal responses were obtained from all of the hearing *Ss* and from six of the 12-year and five of the 7-year deaf *Ss*.

Ss were allowed to keep the marbles they had won and promised to keep the game a secret until everyone had had a chance to play.

RESULTS

Task Performance

The means and standard deviations for Series I and II trials to criterion are presented in Table 1. The learning data for both Series I and Series II were skewed, therefore an analysis of variance was performed for both the raw data and its square root transformation for each series. Both analyses for each series gave an identical pattern of results and the only significant effect was age (Series I, transformed $F = 11.5$, $p < .01$; Series II, transformed $F = 8.5$, $p < .01$).

TABLE 1
MEANS AND STANDARD DEVIATIONS FOR S-I AND S-II TRIALS TO CRITERION (9/10) EXCLUDING CRITERION TRIALS

Group	Age	S-I	S-II
Hearing	7	53.96 (58.33)	19.12 (28.78)
Hearing	12	18.62 (23.01)	7.75 (11.80)
Deaf	7	51.25 (49.52)	22.42 (24.23)
Deaf	12	24.75 (35.53)	9.62 (19.13)

Table 2 presents the proportion of *Ss* in each class: Reversal (R) nonreversal (NR) and inconsistent (In); based on their performance in Series III. There was no significant difference in the proportion of *Ss* choosing a R shift between the deaf and hearing children of either age group, but there was between the two ages ($Z = 2.15$, $p < .032$) (Two-tailed test). An analysis of variance for the number of correct responses in Series III (including both test and nontest pairs) revealed a signifi-

TABLE 2
PERCENTAGE OF *Ss* IN EACH PERFORMANCE CLASS FOR EACH AGE AND CONDITION BASED ON S-III PERFORMANCE

Age	% R	% NR	% In
7	58.3	25.0	16.7
7 (Deaf)	50.0	33.3	16.7
12	71.0	16.7	12.5
12 (Deaf)	79.2	12.5	8.3

cant effect only for age ($p < .05$). These results for the hearing children are quite consistent with those reported by Kendler *et al.* (1962).

Following the analysis used by Kendler and Kendler (1959), all four groups were each divided into fast and slow learners based on their performance in Series I by dividing each group at the median. Among the fast learners there were 35 R out of 48 possible, whereas among the slow learners there were 27 out of 48. There was no significant relationship between speed of learning and the frequency of those choosing a R shift (reversers) (χ^2yates $= 2.23$, $p > .05$).

The presence of a mediator (i.e., a later R shift choice) was not related to the speed of learning Series I. Even if the presence of a mediator cannot be demonstrated to aid in the original learning, this is not as important as the effect it should have in the performance in Series II, where the mediator should give rise to faster shifting (Kendler, 1964). The same analysis was made for Series II data. Here among the fast (below median) shifters there were 40 reversers out of 48 possible, whereas among the slow shifters there were only 22 out of 48 (χ^2yates $=$ 13.07, $p < .001$).

Verbalization

All the hearing children, but only 11 of the 48 deaf children gave a verbal report following completion of the task. Since it is difficult to interpret the verbal report of *Ss* who did not perform consistently in Series III, and since less than a quarter of the deaf children gave any report at all, only the results for the R and NR hearing children will be considered. Table 3 presents the number of *Ss* in each performance and verbalization class. Among the 7-year *Ss*, 70% gave a verbal report which matched their performance in Series III and among the 12-year *Ss*, 76% also matched their report with their performance. There was no significant difference between R and NR *Ss* in the frequencies of correct verbalizations (23/31 vs. 7/10), nor was there a significant relationship between age and verbalization.

TABLE 3
NUMBER OF CONSISTENT *Ss* GIVING VERBAL REPORTS FOR EACH PERFORMANCE CLASS (HEARING CHILDREN ONLY)

	Age[a]				
	7 years		12 years		Combined correct (%)
Verbal Report	Correct	Incorrect	Correct	Incorrect	
Performance					
Reverser	10	4	13	4	74%
Nonreverser	4	2	3	1	70%
Total Correct	14/20 = 70%		16/21 = 76%		

[a] Age vs. verbal reports χ^2yates = 0.09 n.s. Performance vs. verbal reports χ^2yates = 0.02 n.s.

An analysis of the relationship between speed of shifting and verbalization was made for the consistent hearing children, revealing that fast shifters have more correct verbalizations than slow shifters (χ^2yates = 5.82; $p < .025$) (Table 4).

TABLE 4
FAST (BELOW MEDIAN ON S-II) AND SLOW SHIFTERS VS. VERBALIZATION (CONSISTENT HEARING CHILDREN ONLY)

	Fast	Slow	Total
Correct	20	10	30
Incorrest	2	9	11
Total	22	19	41

χ^2yates = 5.82, $p < .025$.

DISCUSSION

There are at least two simple hypotheses which can be examined with the results of this study:

Hypothesis I—Mediation in this paradigm is due to language mechanisms.

Hypothesis II—Mediation is due to some symbolic system (not necessarily verbal) which *is* verbal in hearing children but something else in deaf children.

Despite their general handicap in language skills which is further evidenced by the low number who were able to verbalize at the end of the task, deaf children did about as well as hearing children in Series I,

II and III.[3] This suggests that mediation in this paradigm is not necessarily due to verbal mechanisms, and Hypothesis I can be rejected. However, it is possible that some symbolic response other than ordinary language was operating in the deaf children and that for hearing children the mediation *is* verbal.

On the assumption that verbal mediation, even if covert, would prime the appropriate response systems, it would be expected that children who were mediating via verbal mechanisms would, when asked which card was correct, have a prepotent response of emitting the correct label since there are only four possible simple answers.

If the mediation were due to verbal mechanisms and if the posttask verbalization can be viewed as a good index of the operation of this mechanism then we would expect the relationship between type of shift and verbal report to be: (1) 100% of the R *Ss* verbalize correctly and (2) 50% of the NR *Ss* verbalize correctly (since there are only two concepts being rewarded during Series II). Instead of this outcome, we find: (1) 74% of the R *Ss* gave correct reports and (2) 70% of the NR *Ss* gave correct reports (Table 3).

Since frequency of R shifts increases with age (58 to 71% for hearing *Ss*), a view that this increase is due to verbal mediation would have as a reasonable requirement that correct verbal reports would correspondingly increase with age. The increase of frequency of correct verbalizations with age (70–76%) was not significant ($p > .75$) and was of less than half the magnitude necessary for a good fit.

Since there was no significant difference between hearing *Ss* choosing R or NR in their frequency of correct verbal reports, and since frequency of R shifts increases with age but frequency of correct verbalization does not it appears that Hypothesis II can be rejected, but the data are not that straightforward. For all *Ss* those who were classified as R shifters performed above the median in Series II ($p < .001$) and for consistent hearing *Ss* those who were classified as R shifters gave more correct verbalizations ($p < .025$). This is not an artifact due to subdivisions of the groups since for the 41 consistent hearing *Ss*, those who were classified as R shifters also performed above the median in Series II ($\chi^2_{\text{yates}} = 9.05$, $p < .005$).

This presents a dilemma, for the most direct evidence of the relation-

[3] Only a few of the older deaf children were observed to make any kind of gesture movements with their hands during the task. Deaf children at many modern schools such as those from which these *Ss* came, are *not* taught the manual alphabet, but are encouraged to use whatever "natural" gestures arise in their daily contacts. Thus, the gestures that are used by these older students tend to be large, expressive, and easily observed.

ship between choice of shift and verbalization indicates no relationship exists, yet the indirect evidence of type of shift and verbalization compared to speed of shift indicates a direct relationship should exist. This inconsistency may be resolved by an examination of the definition of mediation.

One problem with the definition of mediation based only on performance in Series III is the lack of an "error term." *Ss* might make a R shift by chance and not be virtue of a mediator. With such a definition, these "pseudo-mediators" are placed in the same class as the "real" mediators. If mediation is at work in this paradigm, one way it should manifest itself would be in facilitating the shift required in Series II, so that *Ss* who were mediating should shift faster than those who were not mediating (Kendler, 1964, p. 428). That this is the case seems well demonstrated in the results (above). However, close inspection of the slow shifters suggests that they are splitting on a chance basis for their performance choices (22:26) and for their verbal reports (10:9) (Table 4). Chance level is difficult to determine for In responders (6 deaf and 7 hearing) but for consistent (R and NR) responders where there are two possible simple concepts being rewarded in Series II, chance would reasonably seem to be 50:50.

In studies of transposition with tasks similar to this, a good argument can be made that the first test trial is really the crucial one since *S* may be learning something during the test series. With this in mind the data were examined for the relationship between first test trial results and classification according to the five test trials. 82 of the 83 consistent performers had their performance on the first trial matching their classification of the five test trials. This provides a basis for classifying those *Ss* who were not consistent for some reason. Of these 13 In, there were 3 "R" and 2 "NR" below and 4 "R" and 4 "NR" above the median on Series II, apparently a chance split between R and NR type responses.

Since slow shifters are apparently choosing a R or NR type of response by chance, it does not seem appropriate to identify mediation in this paradigm by the type of shift alone. The speed of shift data would seem to supply a needed additional criterion for the determination of mediated behavior, thus weeding out the "pseudo-mediators" and allowing a much tighter analysis of what is responsible for the effects found in this paradigm.

Another problem in the assessment of mediation effects lies in the appropriate theoretical comparisons to be made. If R shifts are mediated and NR shifts are not mediated, such comparisons between the two seem appropriate, but when R shifts are compared with other responses (NR shift and In) combined the results are often striking but obscure the distinction between R and NR responses.

The major emphasis on the difference between R and NR shifts by the Kendlers (Kendler and Kendler, 1962; Kendler *et al.*, 1962) has at times been blurred by emphasis upon the R shift alone (Kendler and Kendler, 1966) wherein they have compared R shifts with NR and In combined giving dramatic effects in favor of the R shift. It seems that the crucial argument for their position is based on the difference between R and NR not between R and other responses. Inconsistent responders are very poor in learning, shifting and verbalizing for some unspecified reason. When these *S*s are grouped with the NR *S*s, it then magnifies the relative efficiency of the R *S*s in a way which is not really appropriate to their original theoretical argument.

An example of these effects can be seen by comparing the response choice vs. verbalization analysis for consistent hearing subjects only (Table 3) (χ^2yates $= 0.02$, $p > .75$) with the same analysis which includes the inconsistent responders. For In *S*s there is no easy way to determine if their verbalizations are correct or not, though for most of them it is obvious they are wrong for they often give verbal reports unrelated to the reward pattern in S-II (i.e., "the left one," "the white one" which had not been reinforced at all). Now, by pooling the NR and In groups and assuming that the In *S*s gave incorrect verbal reports, as the majority obviously did, we find the relationship between type of response (R vs. other) and verbalization to be almost significant at the .05 level (χ^2yate $= 3.80$; $\chi^2_{1df} = 3.84$ at .05). Regardless of the outcome of this analysis it would seem to be inappropriate if the major important distinction is between R and NR *S*s. Using a conjunctive definition of mediation based on choice and speed of shift might explain the inconsistencies found above. Table 5 presents the frequency of consistent hearing children found in each choice-speed-verbalization class. Though the small number of *S*s per cell precludes a meaningful statistical analysis, the fact that 90% of the fast R *S*s, whereas only 40% of the slow R *S*s verbalize correctly indicates that verbalization is closely related to the conjunctive definition of mediation.

TABLE 5

PERFORMANCE, SPEED AND VERBALIZATION RESULTS (FREQUENCY OF CONSISTENT HEARING CHILDREN IN EACH CLASS)

	Fast		Slow		
	Correct	Incorrect	Correct	Incorrect	Total
R	19	2	4	6	31
NR	1	0	6	3	10
Total	20	2	10	9	41

Thus, these inconsistencies apparently arise mainly because mediation has been defined by R shift choice *alone*, thereby classifying as mediators a number of slow shifters who apparently chose a R shift by chance.

It should be noted that this close association of fast R shifters and verbalization does not necessarily mean that mediation is verbal for hearing children, it does point strongly in that direction and Hypothesis II can be tentatively accepted.

These data suggest that there is no simple or straightforward relationship between choice of a R shift and verbal mediating mechanisms. The results suggest that a better definition of a mediating child might be a conjunctive one—based both on speed of shift *and* choice of shift. In the light of this analysis, taking a R shift choice as the sole evidence for the presence of mediation seems questionable, whereas a conjunctive definition of mediation does much to clarify the obtained results for the hearing children.

REFERENCES

Ewing, I. R., and Ewing, A. G. *The handicap of deafness*. New York: Longmans, 1938.

Kendler, H. H. Verbal factors in problem solving behavior. *Transactions of the New York Academy of Sciences*, 1954, **16**, 348–353.

Kendler, H. H., and D'Amato, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *Journal of Experimental Psychology*, 1955, **49**, 165–174.

Kendler, H. H., and Kendler, T. S. Effect of verbalization on reversal shifts in children. *Science*, 1961, **134**, 1619–1620.

Kendler, H. H. Problems in problem solving research. *Current Trends in Psychology*, Pittsburgh Univ. Press, 1961.

Kendler, H. H., and Kendler, T. S. Vertical and horizontal processes in problem solving. *Psychological Review*, 1962, **69**, 1–16.

Kendler, T. S., and Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *Journal of Experimental Psychology*, 1959, **58**, 56–60.

Kendler, T. S., Kendler, H. H., and Wells, D. Reversal and nonreversal shifts in nursery school children. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 83–87.

Kendler, T. S., Kendler, H. H., and Learnard, B. Mediated responses to size and brightness as a function of age. *American Journal of Psychology*, 1962, **75**, 571–586.

Kendler, T. S. Development of mediating responses in children. *Monographs of the Society for Research in Child Development*, 1963, **28**, 33–48.

Kendler, T. S. Verbalization and optional reversal shifts among kindergarten children. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 428–436.

Kendler, T. S., and Kendler, H. H. Optional shifts of children as a function of number of training trials on the initial discrimination. *Journal of Experimental Child Psychology*, 1966, **3**, 216–224.

Lenneberg, E. H. *Biological foundations of language*. New York: Wiley, 1967.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 419–433 (1969)

Observing Responses, Visual Preferences, and Habituation to Visual Stimuli in Infants[1]

LESLIE B. COHEN[2]
University of California, Los Angeles

Two experiments investigated the effects of moving visual stimuli on observing responses of 2- to 6-month-old infants. Stimulation consisted of a blinking light which randomly changed position in a 4 × 4 light matrix. In Exp. I, 60 infants were randomly assigned to one of five groups. Four experimental groups were given 24 presentations of a light which either remained stationary or moved among 4, 8, or 16 positions. The fifth group received no light until the last four trials, when they were shown the light moving among 8 positions. In Exp. II, infants' visual preferences were examined by repeatedly presenting two lights simultaneously. Thirty infants were randomly assigned to one of three groups. Group (4, 1) was shown the 4-position and stationary lights, Group (16, 1) the 16-position and stationary lights, and Group (16, 4) the 16- and 4-position lights.

Results from Exps. I and II indicate infants initially fixate longest on stimuli with intermediate position change, and these fixations tend to habituate over trials. Infants were also found to prefer lights which changed position to a stationary light, with the greatest preference occurring on the early trials when the light varied among 4 matrix positions and on the late trials when the light varied among 16 positions. Finally, evidence from Exp. II but not from Exp. I indicates that lights with many position changes produce less habituation than lights with few position changes.

Much research recently has been directed toward the understanding of visual exploration and preference in human infants. The two variables manipulated most frequently have been either the novelty of the stimulus or some aspect of the variability or physical complexity of the stimulus pattern. Most experimental evidence with *Ss* 2 months of age or older indicated the more novel the stimulation, the more infants will observe it

[1] This study is based upon a dissertation submitted to the University of California, Los Angeles in partial fulfillment of the requirements for the doctoral degree. The author expresses his appreciation to Wendell E. Jeffrey for his helpful guidance, to Mary Webster, Tamar Zelniker, and Evelyn Stern for serving as observers in the experiments, and to Kennedy Hill and Dale Kaess for their suggestions and criticisms of this manuscript.

[2] Now at the University of Illinois, Urbana, Illinois.

(Fantz, 1964; Saayman, Ames, and Moffett, 1964). In some instances complex patterns also produce greater fixation times than simple patterns (Berlyne, 1958; Fantz, 1958; Munsinger and Weir, 1967). However, Thomas (1965) reported that when the complexity of the stimuli was first determined by adult ratings, infants preferred a stimulus of intermediate complexity. In addition, both the Thomas study and one by Brennan, Ames, and Moore (1966), in which the number of squares in checkerboard patterns was varied, found the complexity level observed most was a function of age, with older infants preferring more complex stimuli than younger infants.

The Brennan *et al.* and Thomas experiments gave the first direct support to several current theories of exploratory behavior (Berlyne, 1960, 1963; Dember and Earl, 1957; Walker, 1964) which predict an intermediate level of physical complexity will produce maximal exploratory behavior. The theories also predict maximal exploration of stimuli with intermediate novelty. In general, both novelty and physical complexity are subsumed under a broader category such as "psychological complexity" (Walker, 1964), "collative properties" (Berlyne, 1960), or "stimulus change" (Fowler, 1965). Although these theories may differ in some of their terminology and assumptions, they all propose the optimal level of stimulation for an organism will contain an intermediate amount of "stimulus change." If the stimulus change in the environment is beyond this optimal level, the organism will tend to reduce the change. On the other hand, if the stimulus change is lower than the optimal level, he will attempt to increase the change.

It should be possible to test the theories further in human infants by investigating the interaction between novelty and physical complexity. If the physical complexity is great enough to place a stimulus above the optimal level for elicitation of infants' observing responses, then the presentation of that stimulus on several successive trials will reduce the novelty, thus making it a more optimal stimulus. For a stimulus initially at optimal level, the reduction of novelty with successive presentations should reduce it below the optimal level. Finally, for a stimulus below the optimal level, successive presentations should place it even further below the optimal level for eliciting observing responses.

Therefore, the prediction can be made that rapid habituation of infant visual observing responses should accompany stimulation at or below the optimal level, while stimulation beyond the optimal level should result in an early increase in visual observation followed at some later time by a decrease.

Variations in visual preference over trials can also be predicted if one assumes differential habituation reflects preference differences. When a

stimulus of intermediate stimulus change, i.e., near the optimal level, is first paired with one of low change, i.e., below the optimal level, there should be an initial preference for the former. As trials continue, visual exploration to both stimuli should undergo habituation and the initial preference ought either to be maintained or to decrease. When the stimulus of intermediate change is first paired with one of high change, i.e., beyond the optimal level, the initial preference should also be to the intermediate stimulus. However, in this case, since the high-change stimulus should habituate more slowly than the intermediate one, preference ought to be reversed on the later trials. Finally, when the stimulus below the optimal level is paired with one beyond the optimal level, there should be little initial preference for one over the other. As trials continue, the stimulus with high change should come to be preferred over the one with low change.

The present investigation included two experimental tests of these predictions in human infants. Both studies examined the relationship between stimulus novelty and stimulus complexity. Several levels of complexity were produced by varying the number of position changes taken by a single flashing light in a 4×4 light matrix. In Exp. I, each stimulus was shown to a separate group of infants. Fixation times to the matrix stimuli on the first trial were used to estimate the initial optimal level, and changes in responsiveness over trials were examined to see if habituation would be a function not only of repeated presentations of the same stimulus, but also of the degree of complexity associated with the stimulus. In Exp. II, a paired-comparison procedure was used to investigate infants' visual preferences for light-matrix stimulation. The terminology of earlier experiments was followed by assuming an infant preferred a stimulus if he fixated reliably longer on that stimulus than on another simultaneously presented stimulus. The stimuli were the same as in Exp. I, but now two flashing lights which differed in number of position changes were presented concurrently. Both over-all preferences and modification of preferences over trials were examined.

EXPERIMENT I

Method

Subjects. Seventy-three infants were drawn from the Marion Davies Children's Well Baby Clinic at the University of California, Los Angeles. Thirteen had to be replaced because they either fell asleep during the experiment and could not be easily awakened, or they fussed and cried to such an extent that the observers could no longer see their eyes. The final 60 *Ss* were randomly distributed among five groups as

follows: Group E1 (5 boys, 7 girls); Group E4 (8 boys, 4 girls); Group E8 (7 boys, 5 girls); Group E16 (6 boys, 6 girls); and the control group, Group C (4 boys, 8 girls). The infants ranged in age from 8.7 to 23.9 weeks with a mean age of 15.9 weeks.

Apparatus. A standard infant seat and pad, reclining at a 33-degree angle, was placed on the center of an examining table. The seat faced a shelf 18 inches above one edge of the table. Two observers standing at the far edge of the table, 3 feet behind the *S*, viewed the infant's head and eye movements in a small mirror, 7 inches in diameter, which was attached to the center of the shelf. Two types of visual stimulation, a spinning black and white disc and a 4×4 matrix of lights, were also placed on the shelf, one 12 inches to the left of the mirror, the other 12 inches to the right. The position of the two stimuli was reversed for half the *S*s. The stimuli were 40 inches from the infant's eyes and subtended a visual angle of approximately 34 degrees. White sheets covered extraneous and potentially distracting stimuli on the wall and under the shelf.

The light matrix was constructed from a $12 \times 12 \times 3$-inch aluminum box, painted flat black. Sixteen holes, drilled 2 inches apart, formed a 4×4 matrix. Into each hole was inserted one Dialco 18400-935 bulb socket. The socket contained a GE 1829, 28-V bulb covered by a Dialco 95-935 frosted white jewel, $\frac{5}{8}$ inches in diameter. The illumination from each light measured less than one foot-candle at the approximate location of the infant's eyes. Multiconductor cable connected the light matrix to the programming apparatus described below.

The other visual stimulus was a black and white disc, 10 inches in diameter. It was divided into six equal, alternating black and white sections and was rotated by a variable speed motor. The spinning disc was used between trials to turn the infant's head and eyes away from the light matrix.

A 28-V, DC power source supplied the light matrix through a Grason-Stadler Counter. The stepping switch in the counter was moved to a new position by an electronic timer. Snap leads from the matrix were attached to every other position on the counter, and the timer was set to produce a blinking light with both on and off intervals of .2-second duration. By properly arranging the snap leads, one could have either the light blink on and off in the same spot or have it move to any one of the other 15 matrix positions. In order to reduce the noise level of the equipment, the power supply, counter and timer were put into an insulated aluminum ice chest. The ice chest was placed under the examining table approximately 2 feet behind the infant. Four Hunter Timers, also under the examining table, controlled the time the black and white disc spun, the

interval before light-matrix stimulation, the duration of actual light stimulation, and the interval to the start of the next trial.

Each observer held a hand switch which he pressed whenever the infant looked at the light matrix. The switch led to the Driver Amplifier of a Grass Model 5D Polygraph. Because each switch was also in series with a resistor of a different value, the polygraph pen produced a greater excursion for one switch than for the other. With this equipment, it was possible to measure each observer's estimate of infant fixation to the nearest .4 seconds. A second pen recorded the interval of blinking light stimulation and intertrial interval. The Grass Polygraph was located against the far wall of the room, behind and to the left of the infant.

Procedure. Each infant was obtained from the clinic waiting room. He was taken to the examining room and placed in the infant seat while his mother stood near by. The mother then retreated to a chair next to the polygraph where she could watch the experiment but was out of the infant's view. The *E*, who was seated next to the programming equipment, waited for the *S* to look in the direction of the stimuli, then started the first trial. Every trial began with the black and white disc spinning for 3 seconds. Three seconds after the disc stopped rotating, the infant received 12 seconds of light-matrix stimulation. After a final 2-second interval, the next trial began and the disc started spinning once more.

Infants were randomly assigned either to one of four experimental groups or to the control group. The procedure was the same for all groups except for the type of light-matrix stimulation. In the experimental groups, stimulation consisted of a light blinking 32 times on every trial. Between groups, the light differed in the number of position changes it made in the matrix. In Group E1, the light flashed on and off in one position only. In Group E4, the light varied among four possible positions according to a predetermined random order which was 16 blinks long and was repeated twice per trial. Two restrictions were placed on the light's movement: (1) One position came from each row of the matrix; (2) the light never moved to an immediately adjacent matrix position. The last restriction held for Groups E8 and E16 also. In E8, the light moved through eight positions with two positions coming from each row. Finally, in E16, the light moved among all 16 matrix positions.

The control group was split into two halves. Six *S*s (Group C1) were given no light stimulation for 20 trials and then 4 trials with the 8-position light. The remaining six *S*s (Group C2) received 24 trials without light stimulation, followed by four trials with the 8-position light. In the following discussion, any references to the "control group" or "Group C" will mean C1 and C2 combined.

All infants were given a total of 24 trials, with the exception of the

Ss in Group C2, who received 28 trials. The experimental session lasted approximately 15 minutes.

Results

Interobserver reliability was determined by computing a separate Pearson correlation coefficient for each *S* based on 24 pairs of scores. The scores were the total fixation time on each trial during the 12-second interval of light-matrix stimulation. Since the distribution was markedly skewed, each correlation was transformed into a Fisher z' score. The average z' produced an adjusted mean correlation of +.93. This result is consistent with that reported by Cantor and Meyers (1965), who, using similar experimental and statistical procedures, also obtained a correlation of +.93 between observers' ratings of 5-month-old infants' fixation times.

All subsequent fixation-time analyses used the mean of the two observers' scores during light-matrix stimulation. Infants within each group were split at the median into older and younger *Ss*. Their mean ages in weeks are shown in Table 1. Although Group E1 appears to

TABLE 1
MEAN AGES IN WEEKS OF INFANTS IN EACH GROUP

	C	E1	E4	E8	E16	
Old	17.57	21.15	17.38	17.74	18.55	18.48
Young	13.21	14.64	13.50	12.69	12.67	13.34
Mean	15.39	17.90	15.44	15.22	15.61	15.91

contain slightly older *Ss*, the ages of the infants did not differ significantly across groups, F (4, 55) = 1.17. A 5 × 2 (Groups × Age) analysis of variance tested fixation times on the first trial. The only reliable difference was between groups, F (4, 50) = 4.29, $p < .01$. Duncan's Multiple Range Test using the .05 level revealed that *Ss* in Group E4 fixated on the matrix significantly longer than *Ss* in Group E8, Group E16 or Group C. In addition, Group E1 looked at the matrix significantly longer than Group C, but did not differ significantly from E4, E8, or E16. The mean fixation times for control and experimental groups on the first trial are presented in Fig. 1.

Fixation times to the light matrix over blocks of four trials were tested by a 4 × 2 × 6 (Groups × Age × Trial Blocks) analysis of variance with one repeated measure. Subjects in the control group were eliminated from this analysis since six of them received a change in stimulus conditions on trials 21–24. The main effect of trial blocks, F (5, 200) =

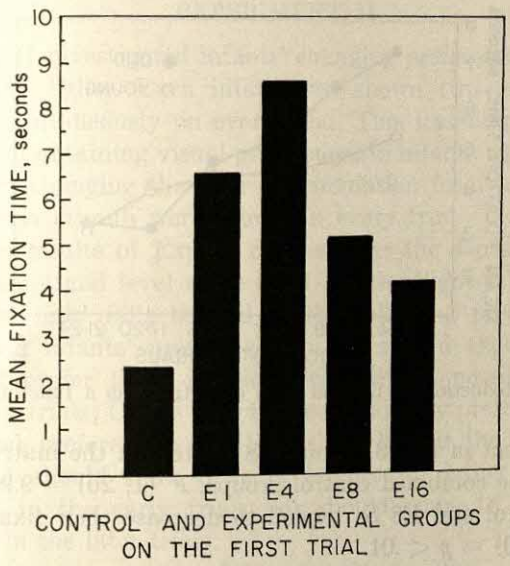


FIG. 1. Fixation times to the light matrix on the first trial for Groups C (no light), E1 (stationary light), E4 (light moving among four positions), E8 (light moving among eight positions), and E16 (light moving among sixteen positions).

19.93, $p < .01$, and the Trial Blocks $\times$ Age interaction, F (5, 200) = 3.82, $p < .01$, were statistically significant. The significant trial blocks effect resulted from the tendency of all groups to decrease their visual fixations over trials.

The absence of a significant Trial Block $\times$ Group interaction indicates there is no statistical justification for assuming one group decreases its fixations over trials more rapidly than another group. The analysis also does not support the hypothesis that the experimental groups differed in overall fixation time.

Figure 2 shows the source of the Trial Blocks $\times$ Age interaction. In the early trials the older infants fixate more than the younger ones. However, the slope of the curve is steeper for the older *Ss*, so that by the later trials, the two age groups fixate on the matrix approximately the same length of time.

Finally, Group E8 was compared with the two control groups, C1 and C2. Group C1 received blinking light which moved among eight matrix positions for the first time on Trial 21. Group C2 was given 8-position stimulation for the first time on Trial 24. Since the control groups were treated identically for the first 20 trials, they were combined and tested against Group E8 on Trials 1–20 by a $2 \times 2 \times 5$ (Groups $\times$ Age $\times$ Trial Blocks) analysis of variance.

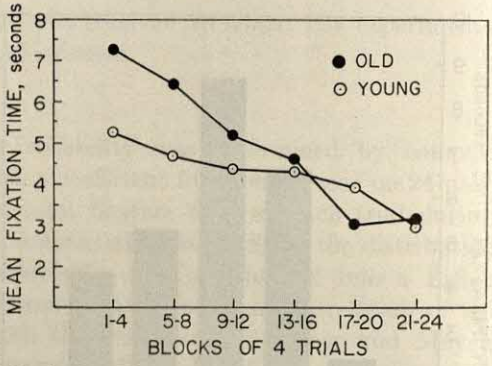


FIG. 2. Reduction in fixation time over trials as a function of age.

As is apparent in Fig. 3, Group E8 fixated on the matrix significantly longer than the combined control groups, F (1, 20) = 9.98, $p < .01$ and both the control and E8 *S*s tended to decrease their fixation time over trials, F (4, 80) = $p < .01$.

During light-stimulation Trials 21–24, Group C1 looked at the matrix significantly more than C2, F (1, 10) = 13.29, $p < .01$. However, as soon as the blinking light was presented to C2, the difference between the control groups disappeared. This is shown by the result that fixation times for C1 on trials 21–24 did not differ reliably from those for C2 on Trials 25–28, F (1, 10) = <1. In fact, once the control groups were given blinking light stimulation, their combined fixation times did not differ significantly from those of Group E8 on Trials 1–4, F (1, 22) = <1.

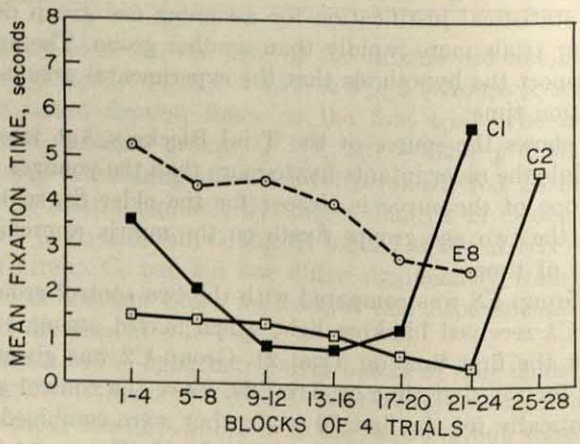


FIG. 3. Comparison of fixation times over trials of the eight-position group with the control groups.

EXPERIMENT II

Experiment II investigated infants' changing preferences for the 1-, 4-, and 16-position lights. Each infant was shown two patterns of light stimulation simultaneously on every trial. This technique resembled the usual method of obtaining visual preferences in infants with the exception that instead of changing the type of stimulation from trial to trial, the same two matrix stimuli were shown on every trial.

If, from the results of Exp. I, one assumes the 4-position light falls at or near the optimal level while the 1-position light falls below it and the 16-position light falls beyond it, the following predictions can be made regarding infants' preferences in the second experiment: Group (4, 1) should prefer the 4-position light with a possible decrease in preference over trials; Group (16, 4) should initially prefer the 4-position light, but switch preference to the 16-position light in the later trials; and Group (16, 1) should show little preference for either the 1- or 16-position light in the early trials but should come to prefer the 16-position light in the later trials.

Method

Subjects. Thirty-seven infants from well-baby clinics in the Los Angeles area were tested in Exp. II. Seven infants had to be replaced because crying forced an early termination of the experiment. The remaining 30 *Ss* were randomly assigned to one of three experimental groups with seven infants in each group coming from the Marion Davies Children's Clinic at U.C.L.A. and three infants coming from the Santa Monica Child Health Conference. The groups each contained five boys and five girls. The infants ranged in age from 11.6 to 24.6 weeks, with a mean age of 17.6 weeks. Within each experimental group, 5 *Ss* were above and 5 below this mean.

Apparatus. The apparatus was similar to that described in Exp. I. The infant seat was on top of a desk facing two identical 4×4 light matrices. Four switches were added to the programming apparatus of the first experiment. By proper use of the switches, each matrix could display a blinking light which remained in one position only, varied among four positions, or varied among 16 positions. The pattern of stimulation was changed from trial to trial through coordination of both sets of switches.

The matrices, one of which had been employed in the first experiment, were placed 3 inches apart on a shelf. They were 24 inches above the desk and approximately 28 inches from the infant's eyes. A 20-inch-wide $\times$ 48-inch-high black masonite board was mounted above the shelf and provided a uniform background for the matrices. Other extraneous stimuli were eliminated by black cloths running from the edges of the board to

the far end of the desk behind the infant seat. Another black cloth hung down from the shelf. Two 1 × ½-inch rectangular peepholes were cut in this latter cloth 9 inches below the shelf. One was located directly between the two matrices; the other was to the infant's right, 1 inch beyond the right-hand matrix.

One observer viewed the infant through the center peephole and recorded visual fixations. With six *S*s, a second observer was also present, recording fixations from the right-hand hole. The observers held two hand switches, one for each light matrix. The pressed the appropriate switch whenever the infant fixated on a matrix and neither observer could see the flashing lights in the matrices. An Esterline Angus Event Recorder indicated observers' responses and the period of stimulation on each trial. Event recorder and programming apparatus were placed on the floor behind the infant seat and desk.

Procedure. Infants were randomly assigned to one of the following groups: Group (4, 1) received simultaneous presentations of the 4-position light and the 1-position light; Group (16, 1) received the 16- and 1-position lights; and Group (16, 4) the 16- and 4-position lights. The experimenter, who was seated next to the programming apparatus, waited until the infant was looking in the direction of the matrices before starting the first trial. Every trial consisted of a 12-second stimulation period followed by a 4-second intertrial interval, and each *S* was given a total of 24 trials. The pattern of stimulation in the two matrices was switched from trial to trial according to a prearranged schedule which required each pattern to be presented three times on each side during every block of six trials. The order was reversed for half the infants in each group.

Results

Interobserver reliability was computed from the six *S*s who had both observers present. As in Exp. I, a separate Pearson correlation coefficient was obtained for each infant, based on 24 pairs of scores. The scores came from observers' estimates of fixation time on each trial during the 12 seconds of matrix stimulation. Since one observer sat to the right of both matrices, his estimates of fixations to the near matrix might be more accurate than to the far matrix. For this reason, two correlations were calculated for every *S*, one for each matrix. Adjusting the means by use of z' transformations produced average correlations of .94 for the near matrix and .90 for the far one.

Fixation times to the patterns of moving light and changes in fixation times over blocks of six trials are indicated for each group in Fig. 4. The data for these curves and for subsequent analyses came from the observer

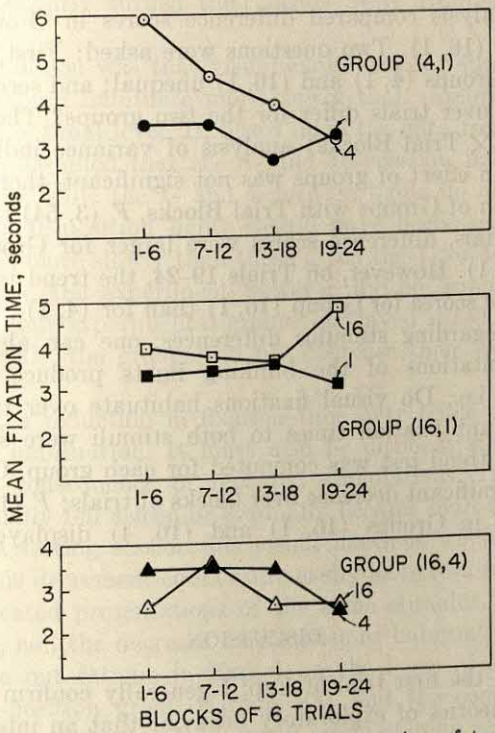


FIG. 4. Changes in fixation time over trials as a function of type of stimulation, plotted separately for Group (4, 1), Group (16, 1), and Group (16, 4).

located between the two matrices. Difference scores between fixation times to the two patterns of stimulation were obtained from every *S*, one score for each block of six trials. The scores in all three groups were computed by subtracting fixation times to the light with fewer position changes from fixation times to the light with more changes, i.e., 4 minus 1, 16 minus 1, and 16 minus 4. In every group, *S*s were divided at the median into older and younger infants.

A separate 2×4 (Age $\times$ Trial Blocks) analysis of variance of difference scores was computed for each group of subjects. The results show that *S*s in Group (4, 1) looked significantly longer at the 4-position light than at the 1-position light, F (1, 8) $= 10.02$, $p < .05$, and Group (16, 1) looked longer at the 16-position light than at the 1-position light, F (1, 8) $= 15.51$, $p < .05$. The difference in fixation times was not reliable in Group (16, 4). Neither the age nor the trial blocks effects reached statistical significance in any group. The general finding, then, was that a light which changes position is fixated more than a stationary light.

The next analysis compared difference scores in Group (4, 1) with those in Group (16, 1). Two questions were asked: First, are the difference scores of Groups (4, 1) and (16, 1) unequal; and second, do changes in these scores over trials differ for the two groups? The results of the 2×2 (Groups $\times$ Trial Blocks) analysis of variance indicated that, although the main effect of groups was not significant, there was a significant interaction of Groups with Trial Blocks, F (3, 54) $= 4.86$, $p < .05$. In the early trials, difference scores were larger for Group (4, 1) than for Group (16, 1). However, on Trials 19–24, the trend is reversed, with larger difference scores for Group (16, 1) than for (4, 1).

Finally, disregarding stimulus differences, one can also ask whether repeated presentations of the blinking lights produce a decrease in fixation times; i.e., Do visual fixations habituate over trials? On each trial block, infant fixation times to both stimuli were summed, and a separate linear trend test was computed for each group. Only Group (4, 1) showed a significant decrease over blocks of trials: F (1, 9) $= p < .05$. Fixation times in Groups (16, 1) and (16, 4) displayed little or no habituation.

DISCUSSION

Results from the first trial of Exp. I generally confirm the prediction from several theories of exploratory behavior that an intermediate level of stimulus change will produce maximal visual exploration. Infants in Group E4, who received the 4-position blinking light, had the longest mean fixation time. Other groups, who received matrix stimulation involving a greater number of position changes (Groups E8 and E16), fixated significantly less. On the other hand, the control group, who on the first trial were not even shown a blinking light, also observed the matrix less than E4. Group E1 is more difficult to interpret. Its mean fixation time fell between Groups E4 and E8 and was not significantly different from either. One can only speculate whether Group E1's true mean lies below that of Group E4.

One must also be somewhat cautious when comparing the control group with the experimental groups. It may be that the added illumination in the room produced by a single blinking light increases the general activity of the infant, and thereby indirectly increases head and eye movements in the direction of the matrix. However, the evidence available on the question does not support this interpretation. Early experiments by Irwin and Weiss (1934), Redfield (1939), and Weiss (1934) showed that as the light intensity increases, the activity level of newborn infants decreases. In one experiment where a bright light did increase activity,

the infants frequently turned their heads away from the light source (Bryan, 1930).

A second prediction was that visual exploration of stimuli beyond the optimal level should habituate more slowly than exploration of stimuli at or below the optimal level. If, based upon the data from the first trial of Exp. I, it is assumed that the 8- and 16-position lights fall beyond the optimal level while the 1- and 4-position lights do not, then infants given the 8- or 16-position lights should have decreased their fixation times over trials more slowly than those given the 1- or 4-position lights. The results of the first experiment, however, did not support this prediction. While all four experimental groups displayed a reduction in visual fixation over trials, the groups did not decrease their fixation times at different rates.

Furthermore, a reduction in fixation time over trials does not necessarily indicate habituation. It must also be demonstrated that the reduction was a consequence of repeated stimulations. If, for example, an infant gradually fell asleep or gradually became more irritable during the course of a testing session, his visual fixations would probably decrease. Since this decrement could quite easily have resulted from factors other than repeated presentations of the same stimulus, one could not, with assurance, call the decrease an instance of habituation.

One can rule out fatigue in Exp. I on the basis of the comparison between the 8-position group (Group E8) and the control group. On Trials 21–24, six members of the control group (Group C1) received the 8-position light for the first time. Their fixation times became significantly longer than the other six control *Ss* (Group C1) who still had not been given any light stimulation. Since all 12 infants had been treated identically on Trials 1–20, one can conclude the increase in Group C1 was the result of light stimulation. This is further verified by the fact that the infants in Group C2 also increased their fixation time when they first saw the 8-position light on Trials 25–28. If Group E8's decrease in fixation time over trials resulted from fatigue, the control *Ss* should have increased only up to the level of E8 on Trials 21–24. Yet, the first four presentations of light stimulation for Groups C1 and C2 produced an increase in fixation time above that of Group E8 on Trials 21–24 and comparable to E8 on Trials 1–4.

In Exp. II, two of the three predictions regarding preference changes over trials were supported. Infants given a stimulus of intermediate change paired with one of low change (Group 4, 1) showed the greatest preference for the 4-position light on the early trials, while those given the high and low change stimuli (Group 16, 1) exhibited no preference initially, but a definite preference for the 16-position light on the later

trials. The failure to find significant differences in Group (16, 4) is puzzling. Perhaps if more trials had been given, the 16-position light would have come to be preferred. When the 16- and 4-position lights were presented to adults, some commented that they had difficulty discriminating between them. The results with infants may have been as predicted if a 32- or 64-position light had been used instead of the 16-position one.

Group (4, 1) also tended to habituate fixation times over trials, while Groups (16, 1) and (16, 4) did not. Since the total over-all stimulation produced by the two blinking lights involved fewer position changes in (4, 1) than in the other groups, there was some support in Exp. II for the hypothesis that stimuli with high change produce less habituation than stimuli with lower change.

Finally, both Exp. I and Exp. II demonstrate the value of light-matrix stimulation in studies of infant exploratory behavior. The moving lights are quite attractive to infants and the stimuli can easily be ordered along a continuum of stimulus change. There is the additional advantage that changes in stimulation are controlled by the experimenter rather than the infant. Fowler (1965) assumes that when an organism scans a complex pattern, he receives more changes in visual stimulation than when he scans a simple pattern. While this may be true for most organisms, an experiment by Salapatek and Kessen (1966) found that when newborn infants were shown a triangle they did not scan the whole figure, but looked primarily at one vertex. If complex and simple patterns are presented simultaneously, one cannot be certain an infant will be getting more changes in stimulation when viewing the complex one. In the experiments above, however, the more the light moved, the more changes in stimulation, even if the infant fixated only on one portion of the matrix.

In conclusion, the present investigation provides some support for current theories of exploratory behavior which assume preference for an intermediate amount of stimulus change. The support is not conclusive, however. Future research must examine the generality of the present results using stimuli other than patterns of moving lights and *Ss* of different ages. Only then can the ultimate utility of such theories be determined.

REFERENCES

BERLYNE, D. E. The influence of the albedo and complexity of stimuli on visual fixation in the human infant. *British Journal of Psychology*, 1958, **49**, 315–318.

BERLYNE, D. E. *Conflict, arousal, and curiosity*. New York: McGraw-Hill, 1960.

BERLYNE, D. E. Motivational problems raised by exploratory and epistemic behavior. In Koch, S. (Ed.), *Psychology: A study of a science*. Vol. 5, Pp. 284–

364. *The process areas, the person, and some applied fields: their place in psychology and in science.* New York: McGraw-Hill, 1963.

BRENNAN, WENDY M., AMES, ELINOR W., AND MOORE, R. W. Age differences in infants' attention to patterns of different complexities. *Science,* 1966, **151,** 354–356.

BRYAN, EDITH S. Variations in the responses of infants during the first ten days of postnatal life. *Child Development,* 1930, **1,** 56–77.

CANTOR, G. N., AND MEYERS, W. J. Inter-observer reliability in judging infants' observing responses. Paper presented at Society of Research in Child Development Meeting, Minneapolis, Minnesota, 1965.

DEMBER, W. N., AND EARL, R. W. Analysis of exploratory, manipulatory, and curiosity behaviors. *Psychological Review,* 1957, **64,** 91–96.

FANTZ, R. L. Pattern vision in young infants. *Psychological Records,* 1058, **8,** 43–47.

FANTZ, R. L. Visual experience in infants: decreased attention to familiar patterns relative to novel ones. *Science,* 1964, **146,** 668–670.

FOWLER, H. *Curiosity and exploratory behavior.* New York: Macmillan, 1965.

IRWIN, O. C., AND WEISS, LABERTA A. Differential variations in the activity and crying of the newborn infant under different intensities of light: a comparison of observational with polygraphic findings. *University of Iowa Studies in Child Welfare,* 1934, **9,** 137–147.

MUNSINGER, H., AND WEIR, M. W. Infants' and young children's preference for complexity. *Journal of Experimental Child Psychology,* 1967, **5,** 69–73.

REDFIELD, JANET. The light sense in newborn infants. *University of Iowa Studies in Child Welfare,* 1939, **16,** 105–145.

SAAYMAN, G., AMES, ELINOR W., AND MOFFETT, ADRIENNE. Response to novelty as an indicator of visual discrimination in the human infant. *Journal of Experimental Child Psychology,* 1964, **1,** 189–198.

SALAPATEK, P., AND KESSEN, W. Visual scanning of triangles by the human newborn. *Journal of Experimental Child Psychology,* 1966, **3,** 155–167.

THOMAS, H. Visual-fixation responses of infants to stimuli of varying complexity. *Child Development,* 1965, **36,** 629–638.

WALKER, E. L. Psychological complexity as a basis for a theory of motivation and choice. In Levine, D. (Ed.), *Nebraska Symposium on Motivation.* Lincoln, Nebraska: University of Nebraska Press, 1964, Pp. 47–95.

WEISS, LABERTA A. Differential variations in the amount of activity of newborn infants under continuous light and sound stimulation. *University of Iowa Studies in Child Welfare,* 1934, **9,** 1–74.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 434–439 (1969)

The Effects of Verbalization on Verbal Discrimination Learning and Associative Recall in Young Children and Adults[1]

L. R. GOULET AND WILLIAM J. HOYER
West Virginia University

Groups of fifth-grade and college *Ss* learned a verbal-discrimination list consisting of 12-paired high-frequency words. The lists were learned under conditions where *Ss* were or were not instructed to pronounce the right and wrong items during the feedback exposure period. Following acquisition, *Ss* were given eight trials involving recall of Right-Wrong and Wrong-Right associations.

Rate of acquisition was directly related to age and pronunciation during acquisition hindered learning for both fifth-grade and college *Ss*. Associative recall of both W-R and R-W associations was facilitated under the pronunciation treatment. However, W-R recall was greater than R-W recall for the college *Ss*, with the reverse being true for the fifth-grade *Ss*. The results were discussed in relation to the frequency theory of verbal-discrimination learning.

Overt verbalization or pronunciation of task items has generally been regarded as having a facilitative effect on children's discrimination learning. Empirical support for this generalization has been found in tasks where children have been given pretraining on naming task stimuli (e.g., Norcross and Spiker, 1957), and where the children are instructed to verbalize the stimulus items during acquisition of a discrimination task (Weir and Stevenson, 1959).

These results are in apparent contradiction to a recent theoretical analysis of verbal discrimination (VD) learning (Ekstrand, Wallace, and Underwood, 1966) formulated to account for data on VD learning in adult *Ss*. In short, frequency theory suggests that acquisition of single VD lists is, in part, a function of the frequency with which right (R) and wrong (W) items in a VD pair have been overtly or covertly verbalized. The theory assumes that the exposure of the R and W items

[1]The writers express their appreciation to Mr. William Cunningham, Principal, Monongalia County Public Schools and the teachers of First Ward School for their helpful assistance in the conduct of this research. This research was supported by USPHS grant MH 13515-02 to the senior author.

of a pair adds a frequency unit to each, but that the R item accrues additional frequency units during the feedback exposure. Additional frequency units are added to the R items if rehearsal occurs. Learning is assumed to occur because *S*s use the differential frequency of R and W items as discriminative cues.

If *S*s are instructed to verbalize R and W items during the feedback exposures of the pairs in the VD list, frequency units would be added to both R and W items, preventing an inordinate increase of frequency units to the R items relative to the W items, with a concomitant detrimental effect on VD learning.

The present experiment involved a comparison of VD learning under conditions where *S*s were or were not instructed to verbalize R and W items during acquisition. Kausler and Sardello (1967) and Sardello and Kausler (1967) have found detrimental effects of pronunciation when adult *S*s were used but no data are available from children. In addition, the successive and contiguous presentation of the R and W items from trial to trial does lead to the learning of R-W and W-R associations, at least with college *S*s (Sardello and Kausler, 1967). Sardello and Kausler found equal amounts of W-R and R-W recall under both verbalization and nonverbalization conditions. Again, however, no parallel data from children are available.

METHOD

Design and Subjects

The *S*s were 32 fifth-grade children and 32 college undergraduates. The children were randomly chosen from fifth-grade of a Morgantown public school while the college *S*s were taken from those enrolled in general psychology classes at West Virginia University. All of the *S*s were naive with respect to verbal-learning tasks.

The *S*s at each age level were randomly assigned to one of eight treatments with the restriction of equal *N*s for each block of eight *S*s. The between-group variables were verbalization (pronunciation of R and W items during the feedback exposure or no overt pronunciation) and two sets of two VD lists (discussed below) constructed to extend generalization of results and to control for item and pair difficulty.

Materials

A pool of 40 associatively unrelated words was selected from the Palermo and Jenkins (1964) word norms. Words of less than three or more than eight letters were excluded. The words were then randomly paired to form two different VD lists of 10 pairs each. No pair contained

words beginning with the same letter of the alphabet. The correct word in each pair was also selected randomly, and four different serial orders of the list were constructed for each list to control for serial position cues. In addition, the correct item of each pair occurred in the right and left position twice in the four serial orders, with the restriction that five R items were presented in the left position in each order.

Procedure

The VD pairs were presented on a Stowe memory drum at a 2:2-second rate with a 4-second intertrial interval. First the pair appeared, then the drum turned, and the pair appeared again, this time with the correct item underlined. Each *S* was given standard VD instructions. In addition, the *S*s in the pronunciation (P) treatment were asked to pronounce the items in the pair once each during the feedback exposure while those in the nonpronunciation (NP) treatment were not instructed to do so. This procedure closely followed that used by Kausler and Sardello (1967).

Each *S* was taken to a criterion of two successive errorless trials on the VD list which was immediately followed by eight trials involving associative recall of both W-R and R-W associations. During the associative recall trials the *S*s were shown either the R or W item of the pair and were asked to recall the W (or R) item with which it had been paired during acquisition.

Two different lists were constructed for associative recall. For the first list, five of the W items were randomly selected to serve as stimuli for W-R recall with the R items from the remaining five pairs serving as stimuli for R-W recall. For the second list the direction of recall was reversed. This procedure permitted measurement of both W-R and R-W recall for each *S* while also controlling for item or pair difficulty. The items were presented at a 2-second rate in four serial orders during the associative recall trials and no further pairings of R and W items were presented.

RESULTS AND DISCUSSION

VD Learning

List differences were unrelated to the dependent variables and were eliminated from all statistical analyses reported below. The mean trials to criterion for the college and fifth-grade *S*s in the P and NP treatments were: College, P: 15.94, NP: 9.69; fifth-grade, P: 26.00, NP: 17.69. Statistically significant main effects for both age and verbalization were found, Fs (1,60) = 12.38; 8.05, $p < .01$, respectively. Age and

rate of learning were directly related and overt verbalization of R and W items during learning hindered the performance of both the young and older *Ss*. The Age × Verbalization interaction did not approach significance ($F < 1$).

The data for VD learning clearly indicate the interfering effects of overt verbalization in VD learning for both children and adults. These effects closely parallel those obtained by Kausler and Sardello (1967) and Sardello and Kausler (1967) who used a VD task and college students as *Ss*, and also those obtained by Carmean and Weir (1967) who used a 10-pair discrimination list with paired-pictures as stimuli.

The problem remains to explain the facilitative effect of verbalization obtained by Weir and Stevenson (1959). In their experiment, *Ss* were required to pronounce the stimulus of their choice for each pair during acquisition. This procedure effectively results in verbalization of the R item of each pair with this effect increasing as performance exceeds the baseline for guessing or chance level (e.g., Ekstrand *et al.*). Carmean and Weir, using college *Ss*, did find that pronunciation of R items alone did facilitate discrimination learning. The present results and those of Carmean and Weir (1967) and Weir and Stevenson (1959), are in accord with frequency theory and suggest that the theory may account for performance differences in VD learning in children as well as adults.

Associative Recall

Table 1 presents summary data for W-R and R-W recall for each treatment and at each recall trial. Consistent with the results of Sardello and Kausler (1967), verbalization facilitated the recall of both W-R and R-W associations, F (1,60) = 11.82, $p < .01$. There were no statistically significant age ($F < 1$), Recall-Direction ($F = 1.91$), nor Trials ($F = 1.25$) main effects. However, a statistically significant Verbalization × Recall-Direction interaction was obtained, F (1,60) = 4.62, $p < .05$. This interaction may be attributed to the disproportionate facilitation of R-W recall relative to W-R recall under the P treatment. This effect was most apparent for the fifth-grade children, i.e., W-R recall was greater than R-W recall for the college students under both P and NP treatments, and for the fifth-grade *Ss* under the NP treatment. However, R-W recall slightly exceeded W-R recall for the children under the P treatment. The Age × Recall-Direction × Verbalization interaction however, did not approach statistical significance, ($F < 1$). The Age × Trials interaction was statistically significant, F (7,420) = 3.65, $p < .01$. As may be seen from Table 1, the college *Ss* are initially better in recall than the children. However, the performance of children increased over the eight trials, while that for the college *Ss* showed a slight

TABLE 1
MEAN W-R AND R-W RECALL FOR EIGHT TRIALS

	Fifth grade: Trials							
	1	2	3	4	5	6	7	8
Pronunciation								
R-W	2.12	2.37	2.44	2.56	2.62	2.56	2.69	2.44
W-R	1.56	1.69	1.87	1.87	1.87	1.94	2.31	1.87
Nonpronunciation								
R-W	1.06	1.56	1.31	1.37	1.37	1.31	1.37	1.31
W-R	1.19	1.25	1.62	1.44	1.69	1.50	1.56	1.31
	College: Trials							
	1	2	3	4	5	6	7	8
Pronunciation								
R-W	2.31	2.25	2.37	2.44	2.25	2.25	2.44	2.37
W-R	3.00	2.88	2.56	2.62	2.62	2.75	2.88	2.75
Nonpronunciation								
R-W	1.06	.88	.88	.94	1.00	1.00	.88	.88
W-R	1.75	1.75	1.62	1.62	1.56	1.75	1.50	1.69

decrease. It is likely that the differences in recall at Trial 1 are due to differences in performance rather than to differences in degree of learning. That is, the performance of children was probably disrupted on Trial 1 by the relatively fast, paced-rate of the recall task. No other interactions exceeded 1.06 in this analysis.

The data relating to associative recall are somewhat in conflict with those of Sardello and Kausler (1967). These authors found equal recall of W-R and R-W associations under both P and NP conditions while the present data indicated greater W-R than R-W recall under these treatments. These differences may be attributed to the paced conditions used in the present study as compared to the unpaced recall period used by Sardello and Kausler.

Finally, Spearman *rho* correlations were computed separately between the number of trials in acquisition and total associative recall over the eight trials for each of the age and verbalization treatments. None of these correlations exceeded .34. However, even with the lack of statistically significant correlations the direct comparisons of associative recall between ages should be interpreted with caution because of the varying numbers of trials to criterion among groups in acquisition. The primary concern here is with the relative strengths of R-W and W-R associations

as a function of the verbalization treatments and the differential effects for each age group.

REFERENCES

CARMEAN, S. L., AND WEIR, M. W. Effects of verbalizations on discrimination learning and retention. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 545–550.

EKSTRAND, B. R., WALLACE, W. P., AND UNDERWOOD, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, 73, 566–578.

KAUSLER, D. H., AND SARDELLO, R. J. Item recall in verbal discrimination learning as related to pronunciation and degree of practice. *Psychonomic Science*, 1967, 7, 285–286.

NORCROSS, K. J., AND SPIKER, C. C. The effects of type of stimulus pretraining on discrimination performance in preschool children. *Child Development*, 1957, 28, 79–84.

PALERMO, D. S., AND JENKINS, J. J. *Word association norms: Grade school through college*. Minneapolis: University of Minnesota Press, 1964.

SARDELLO, R. J., AND KAUSLER, D. H. Associative recall in verbal-discrimination learning as related to pronunciation and degree of practice. *Psychonomic Science*, 1967, 8, 253–254.

WEIR, M. W., AND STEVENSON, H. W. The effect of verbalization in children's learning as a function of chronological age. *Child Development*, 1959, 30, 143–149.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 440–447 (1969)

Adults as Discriminative Stimuli for Different Reinforcement Contingencies with Retarded Children[1]

WILLIAM H. REDD AND J. S. BIRNBRAUER
University of North Carolina at Chapel Hill

One adult dispensed food and praise for play behavior, while a second adult reinforced the retarded children in a group setting on a noncontingent fixed interval schedule. Two children were selected as targets for individual analysis. The entry of the contingent adult evoked play; the noncontingent adult did not influence behavior. After extinction, the adults reversed roles. Conditioning was replicated, supporting the conclusion that contingent reinforcement outweighed the other differences between the two adults (sex, familiarity, status, etc.). Four subsequent tests of one *S* were conducted to evaluate the generality and strength of control. The adults acquired discriminative properties when paired with specific reinforcement contingencies.

Baumrind (1966) has specified three child rearing roles that parents typically assume (authoritarian, authoritative, or permissive) and has enumerated child behavior which is associated with each role. Parents have also been characterized by the form of discipline they employ. Sears, Maccoby, and Levin (1957) have similarly suggested that the form of discipline the parents employ is correlated with particular gross behavior patterns of the child. Thus, one might expect a child's behavior to be in part a function of the disciplinary routine that his parents follow. If his parents do not follow the same routine, he should show discrimination between them. The disobedient child who is a terror for his mother may become quiet and manageable when his father returns home from work. The adults could be considered as cues or signals for the child that certain disciplinary rules or contingencies are in effect.

[1] This research was supported in part by NIMH Project Grant MH 14914. The authors express their appreciation to Ann M. Gardner, Eileen G. Grennan, and Carol A. Trageser for their assistance in conducting this research. This study would not have been possible without the cooperation of Julia Lawler and the boys' cottage parents.

Thus, the child comes to respond differentially to adults depending upon the rules that the adults have enforced in the past.

McCoy and Zigler (1965) as well as Berkowitz and Zigler (1965) have examined the effectiveness of social reinforcement within a given experimental situation as a function of prior adult-child interaction. The general conclusion is that the familiar adult with whom the child has experienced positive interaction is the more effective agent of reinforcement. Socioeconomic status (Sgan, 1967), sex, and race (Allen, Dubanoski, and Stevenson, 1966) of both the child and the agent of reinforcement have also been found to be important determinants of the agent's effectiveness.

The parent's role as an agent of both tangible and social reinforcement has been discussed by Bijou and Baer (1965). They suggest that aspects of the parent's behavior, as well as the parent himself, become not only conditioned reinforcers for the child but also cues or discriminative stimuli. The adult comes to control the behavior of the children in accordance with the particular reinforcers and regimes of reinforcement he employs. One would expect his entrance or exit to "turn on" or "turn off" the children's responding. In order to investigate the stimulus control that adults might acquire through association with certain behavior requirements, we paired adults with different contingencies and compared the children's behavior when these adults were present. An adult who required the child to emit a particular response in order to receive candy and praise was compared with an adult who dispensed candy and praise without regard to their behavior. Hart, Reynolds, Baer, Brawley, and Harris (1968) found that adult social reinforcement of cooperative play of a 5-year-old girl in a preschool setting resulted in a significant increase in cooperative play only when reinforcers were dispensed contingently. Noncontingent reinforcement (reinforcement without regard to *S*'s behavior) did not result in any change over baseline. In their study, the same adults functioned as agents of both contingent and noncontingent reinforcement. After baseline data was obtained, all adults reinforced noncontingently for 7 consecutive days and then contingently for 12 days. If, however, certain adults had always reinforced contingently and others always noncontingently, we wondered how the children would come to respond to these adults. Would responding change when these adults entered and left the room?

A multiple baseline design (Ferster and Skinner, 1957) was employed in this study. All experimental conditions were included in each session, rather than the more common design in which baseline data and experimental data are collected during different days.

METHOD

Subjects

Ten severely retarded boys, residents of Murdoch Center, Butner, North Carolina, were divided into two play groups. CA's ranged from 12–3 to 15–9 and MA's from 2–10 to 4–8. All boys were selected from one ward and assigned to groups based on similar capability and behavior in a free play situation. One boy from each group was selected as a target subject for individual analysis. *S*1 (CA 15–9 and MA 3–6) had a diagnosis of mental retardation associated with psychotic or major personality disorder. His social and play repertoires were extremely limited. He spent most of his free play time on the ward sitting alone on the floor, refusing to play or interact with other children. *S*2 (CA 14–3 and MA 4–4) had the same diagnosis as *S*1, but schizophrenia was specified. On the ward he usually engaged in simple tasks or solitary play, avoiding any interaction with other children. The target *S*s were selected because they showed stable patterns of responding in the experimental playroom during preliminary sessions.

Setting and Procedure

The study was conducted in a 16 × 25-foot room which was equipped with selected toys (balls, blocks, skate boards, and trucks). These toys were always present in the room. Across one corner of the room a partition with a one-way vision screen provided an observation booth.

Throughout the experiment two observers stationed behind the one-way vision screen recorded the behavior of the target *S* once every 2½ seconds. Since the repertoires of the target *S*s were very limited, recording was simple. The observers recorded whether *S*1 was sitting or standing and "playing," and they recorded if *S*2 were engaging in "cooperative play." The criterion for standing and "playing" was that *S* was on his feet and moving a toy. Cooperative play was defined as moving or manipulating a toy with another child. After the observers were trained, the inter-rater percentage of agreement was computed on four separate occasions. Calculations were based on comparisons of 324 observations, and 94 percent was the lowest value obtained.

Each session was divided into three conditions—baseline, contingent reinforcement, and noncontingent reinforcement. The procedure for both target *S*s is summarized in Table 1. The response selected for experimental manipulation was the only difference in the treatment of the target *S*s.

At approximately the same time 4 days a week, the target *S* and four other boys from his ward were taken to the playroom by the two

TABLE 1
OUTLINE OF EXPERIMENTAL SESSIONS

Period	Procedure	Minutes
Baseline	Five *S*s present in the playroom with no adult.	5
Contingent	Adult I dispensed reinforcers contingent upon *S*'s behavior on a FI 45 seconds with a limited hold of 15 seconds.	5
Baseline	Five *S*s present with no adult.	2
Noncontingent	Adult II dispensed reinforcers once every 60 seconds without regard to *S*'s behavior.	5
Baseline	Five *S*s present with no adult.	2
Contingent	Adult I dispensed reinforcers contingent upon *S*'s behavior on a FI 45 seconds with a limited hold of 15 seconds.	5
Baseline	Five *S*s present with no adult.	5
Noncontingent	Adult II dispensed reinforcers once every 60 seconds without regard to *S*'s behavior.	5
Baseline	Five *S*s present with no adult.	5

experimenters. The boys were left alone in the room for 5 minutes. Then, either Adult I (a female) or Adult II (a male) entered the room with a reinforcer cup and a stopwatch in his hand and followed the prescribed routine. In order to check for possible sequence effects, the order in which the adults entered the room was varied. Adult I, the contingent adult, dispensed reinforcers to each *S* only when he was engaging in the target behavior. Adult II, the noncontingent adult, reinforced *S*s without regard to the children's behavior. While in the playroom the adults moved about quietly and only spoke to the *S*s when reinforcing them. Both adults used the same edible reinforcers (M & M's, bites of ice cream, and sips of coke) and said "good boy," "that's right," etc. with every edible given. During the beginning conditioning sessions the contingent adult prompted and reinforced approximations of the target behavior and reinforced the other children on a FI 45 with a limited hold of 15 seconds. The prompts were faded out and FI reinforcement was programed. The interval was lengthened gradually until the FI 45 was reached. Adult II matched the number of reinforcers that Adult I dispensed. When Adult I moved to the FI 45, Adult II dispensed reinforcers once every 60 seconds. Thus, the adults dispensed the same number of reinforcers; the only difference was that Adult I's reinforcers

were dispensed contingently and Adult II's were dispensed noncontingently.

After *S* responded differently to the adults' entries, i.e., emitted the target behavior within the first 45 seconds, etc., for three successive sessions, extinction began. During extinction each adult entered the room on schedule but withheld reinforcement from the target *S*. After stable responding was established for three consecutive extinction sessions, the adults reversed roles. During reversal sessions, Adult I dispensed reinforcers without regard to *S*'s behavior, and Adult II dispensed reinforcers contingently.

RESULTS

The criterion for conditioning, i.e., differential response to the adults' entries for the three successive sessions, was met by *S*1 in 16 sessions and by *S*2 in 28 sessions. The cumulative records of these sessions, plotted in Figs. 1 and 2, show that each *S* emitted the target behavior

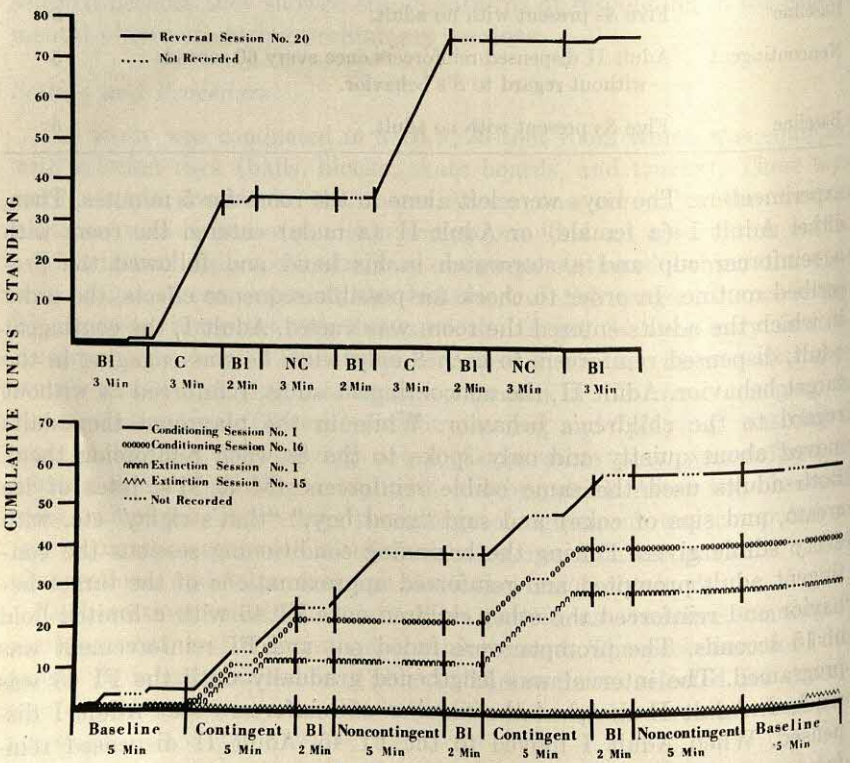


FIG. 1. Criterion data for *S*1 obtained during each phase of the study. Data was not recorded during observer breaks.

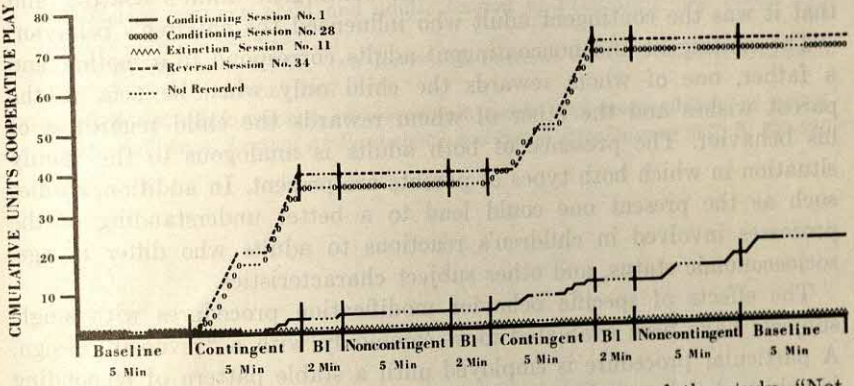


FIG. 2. Criterion data for *S*2 obtained during each phase of the study. "Not recorded" refers to 1-minute observer breaks.

only when the contingent adult was present and that this behavior was controlled by the adult's entry and departure. During the first conditioning sessions the target behavior was much less predictable and occurred considerably less frequently in the case of *S*2. With extinction programed, the presence of either adult had no effect. Conditioning was replicated with the adults in reversed roles, supporting the conclusion that contingent reinforcement outweighed the other differences between the two adults (sex, familiarity, status, etc.).

Sequence of periods had no effect. Several subsequent tests of *S*2 were conducted. (a) Five other adults entered with the stopwatch and reinforcement cup. None evoked cooperative play despite their reinforcement of the other children in the room. (b) The contingent adult entered with the noncontingent adult. (It was as if the noncontingent adult were not present.) *S*2 played cooperatively. (c) The experimental sessions were conducted on *S*2's ward. The change in locale had no effect. (d) The contingent adult entered the playroom without the reinforcement cup. The appropriate behavior did not occur. Control of the target behavior was specific to the currently contingent adult, provided that he was carrying the reinforcement cup.

DISCUSSION

The experiment demonstrated that an adult who dispensed reinforcers contingently acquired discriminative properties and functioned as a cue which influenced the play behavior of the children. The children did not change their behavior when the noncontingent adult entered, and, in fact, they ignored him. When both adults entered the playroom together, the target children played appropriately. This showed that the non-

contingent adult had no effect on the contingent adult's control, and that it was the contingent adult who influenced the children's behavior.

The contingent and noncontingent adults correspond to a mother and a father, one of whom rewards the child only when he acts as the parent wishes and the other of whom rewards the child regardless of his behavior. The presence of both adults is analogous to the family situation in which both types of parents are present. In addition, studies such as the present one could lead to a better understanding of the processes involved in children's reactions to adults who differ in age, socioeconomic status, and other subject characteristics.

The effects of specific behavior modification procedures with single subjects have been evaluated most commonly with a "reversal" design. A particular procedure is employed until a stable pattern of responding is achieved, followed by a second procedure, etc. If the behavior waxes and wanes in correspondence with the conditions imposed, then the procedure, not the passage of time, Hawthorne effect, etc., is credited with the outcome. In the study reported here, the utility of employing a second design, multiple schedule baseline, has been demonstrated.

The multiple schedule baseline method is more reliable and economical than the reversal approach. Within every experimental session control and experimental measures are recorded by the same observers in close temporal proximity. Each session is divided into periods such that each session consists of all phases of the experimental manipulation. The potentially confounding effects of extra-laboratory events and of passage of time on experimenter and observer consistency are reduced. Thus, changes in *S*'s behavior can be more readily attributed to intra-session manipulations than to extraneous variables.

REFERENCES

Allen, S., Dubanoski, R. A., and Stevenson, H. W. Children's performance as a function of race of *E*, race of *S* and type of verbal reinforcement. *Journal of Experimental Child Psychology*, 1966, **4**, 248–256.

Baumrind, D. Effects of authoritative parental control on child behavior. *Child Development*, 1966, **37**, 887–907.

Berkowitz, H., and Zigler, E. Effects of preliminary positive and negative interactions and delay conditions on children's responsiveness to social reinforcement. *Journal of Personality and Social Psychology*, 1965, **2**, 500–505.

Bijou, S. W., and Baer, D. M. *Child development II: The universal stage of infancy*. New York: Appleton, 1965.

Ferster, C. B., and Skinner, B. F. *Schedules of reinforcement*. New York: Appleton, 1957.

Hart, B. M., Reynolds, N. J., Baer, D. M., Brawley, E. R., and Harris, F. R. Effect of contingent and noncontingent social reinforcement on the cooperative play of a preschool child. *Journal of Applied Behavior Analysis*, 1968, **1**, 73–76.

McCoy, N., and Zigler, E. Social reinforcement effectiveness as a function of the relationship between child and adult. *Journal of Personality and Social Psychology*, 1965, **1**, 604–612.

Sears, R. R., Maccoby, E. E., and Levin, H. *Patterns of child rearing*. Evanston, Illinois: Row, Peterson, 1957.

Sgan, M. Social reinforcement, socioeconomic status, and susceptibility to experimenter influence. *Journal of Personality and Social Psychology*, 1967, **5**, 202–210.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 448–458 (1969)

Recognition Memory for Pictured and Named Objects

GRAHAM M. DAVIES[1]
University of Aberdeen, Scotland

Kurtz and Hovland (1953) and others have shown that children who name objects at presentation show superior retention for the material relative to *Ss* who do not. The conventional explanation for this effect is that it reflects the superior potency of the word as a carrier of sense impression compared to nonverbal forms. In the present experiment, type of stimulus (picture/name) and type of response (find picture/name) were systematically manipulated and the effects on subsequent recognition performance ascertained. Results suggested that, for *Ss* in the age range tested (8½–10½ years), a version of the trace modality hypothesis, (Wallach and Averbach, 1955) may provide a more satisfactory interpretation of this phenomenon, than a theory based on verbal potency.

Investigations into the role of naming in the discrimination and retention of perceptual stimuli by children, have generally utilized nonsense shapes rather than meaningful objects (e.g., Spiker, 1956; Katz, 1963). "Meaningful objects" may be defined as stimuli with which the *S* is familiar, and for which an appropriate verbal label is readily available. While the interpretation of the facilitatory effects on the retention of nonsense shapes following labeling practice, is a matter of considerable dispute, the critics are more united in their interpretation of similar results with meaningful material. Naming an object, it is argued, causes it to be lodged in a superior coding system which preserves it from the rapid decay of the visual "image."

Evidence for such a viewpoint is derived from two sources. Firstly there are the unsystematic and anecdotal observations of early researchers, who noted that when *Ss* were asked to memorize the details of a perceptual array, those *Ss* who later reported verbalizing the main features to themselves at the time of observation, tended to retain the

[1] This paper describes work carried out at the University of Hull where the author was in receipt of a Post-Graduate Research Studentship awarded by the Social Science Research Council. I wish to thank Prof. A. D. B. Clarke and Dr. A. M. Clarke for valuable assistance and guidance at all stages of this investigation. I also wish to acknowledge the generous research facilities afforded me by Mr. T. Clark, Chief Education Officer for the East Riding of Yorkshire, the co-operation of the staff and pupils of Brough and Willerby primary schools and the technical assistance of Mr. B. Marshall and Mr. J. Devine.

material better than *S*s who claimed to rely on visual imagery and the like (Kuhlmann, 1907; Bartlett, 1932; Jenkin, 1935). A second source for such a view is an experiment by Kurtz and Hovland (1953). In this study, children aged 10–13 years, were successively exposed to a series of common objects. Half the *S*s were required to circle the name of each object in a list of alternatives, while for the rest of the *S*s, the required response was to circle a picture of each object, on a photograph of all the objects grouped together. In addition, the *S*s who found the names of the objects, said each name aloud as they circled it. A week later, the children were unexpectedly tested for their knowledge of the items presented. In terms of simple recall, the *S*s who had named the objects remembered a significantly greater number of items and gave significantly fewer instrusions than *S*s who had matched each object with a picture. The children were also given a recognition test in which half the material appeared as pictures and half as names. Although *S*s in the object-naming group made significantly fewer errors on this task than the other *S*s, they failed to score significantly higher in terms of total number of items correctly identified, and actually recognized insignificantly fewer items on the picture portion of the recognition test. Despite the somewhat equivocal nature of the results and some ambiguities of design, this experiment has been frequently hailed as demonstrating: "The superior potency of a *word* as a carrier of sense impression . . . even when the experiment is designed so that an equal amount of attention is paid to the stimuli in initial learning" (Carroll, 1964, pp. 95, his italics), or again as ". . . fairly good evidence that verbalisation facilitates retention over fairly long intervals more than visualisation does" (Reese, 1965, p. 291).

These interpretations seem based on two suppositions, firstly, that the advantage of the naming group lay in the possession by *S*s of a list of names as such for the objects presented, and secondly, that the photograph-matching procedure led to the laying down of nonverbal "visual images," while naming did not. This explanation will be hereafter referred to as the verbal potency hypothesis.

However, these assumptions seem somewhat premature in view of the failure of the design to discriminate between the direct and indirect effects of naming. This problem has been seen as increasingly important since the publication of the experiment (see, for instance, Erickson, 1967, pp. 121–142). Kurtz and Hovland themselves put forward two hypotheses, both of which rely on the indirect benefits of labeling to retention. One suggestion was that the result might be a function of the overt nature of the naming response. It was, perhaps, unfortunate, that in their design, the authors failed to control for the effects of giving an object

a name as such, and the effects of repeating that name aloud. Their alternative suggestion was that naming an object served to direct the *Ss* attention toward more distinctive features of each stimulus, than did matching each item with its likeness. Successful naming required precise identification of each object, whereas matching an object to a picture could be achieved purely on the basis of size, shape, shading, or general conceptual category. However, as Wallach and Averbach (1955) pointed out, this latter theory should predict that the picture-matching condition should give inferior recognition scores to the naming condition on the picture section of the recognition test, as the more general cues which sufficed for the matching task could only lead to interference in the test situation. As mentioned previously, this predicted inferiority was not present.

After an extensive review of the experiment, Wallach and Averbach put forward a third suggestion, based on Gestalt theories of memory. They argued that memories are not retained in a common intersensory store but are specific to the modalities in which they are received. Wallach and Averbach were primarily concerned with sensory modalities (e.g., auditory, visual) but in order to cope with Kurtz and Hovland's data, they postulated the existence of symbolic modalities, specifically, visual nonverbal and verbal. They suggested that naming the objects led to the laying down of both visual and verbal traces of the event, whereas matching an object with a picture left only a visual trace. Recall could be considered a simple function of the number of trace modalities in which an item was encoded. Recognition, on the other hand, could take place only if the form of one of the stimulus traces encoded by the *S* matched the mode of presentation of the recognition material. Thus the picture-matching group would perform on equal terms with object-naming group on the picture section of the recognition test, but would be handicapped on the name section through lack of available verbal traces. The only verbal traces available to picture-matching *Ss* would be those which arose through chance overt verbalization during the presentation task and hasty conversions from visual to verbal traces during the actual recognition session. Wallach and Averbach's explanation (hereafter referred to as the trace modality hypothesis) fits the data very well, but failed to arouse experimental interest at the time of publication.

Since the appearance of these reports neither the modality redundancy, nor the verbal potency hypotheses, seem to have been satisfactorily tested, least of all in the context for which they were originally formulated, that is with children memorizing familiar objects. The present experiment uses a modification of the Kurtz and Hovland paradigm, in an at-

tempt to discriminate between the direct benefits to retention derived from possessing an item encoded as a name as such, and any indirect benefits springing from the act of naming, such as trace modality theory suggests.

Four conditions of presentation were used. In the *name-matching* condition (referred to as NN), *S*s were shown a series of names of objects and were required to find each name in a list of names. In the *picture-matching* condition (PP), *S*s were shown a set of pictures of objects and were required to locate each object on a large photograph of all the objects. For the *picture-naming* condition (PN), the same set of photographs were used, but this time the *S* found the names of each of the items shown. The reverse of this was the *name-picturing* condition (NP), where *S*s were shown the names of the objects and had to find each object in the large group photograph. The PP and PN conditions were similar to Kurtz and Hovland's picture matching, and object-naming groups, respectively.

It was predicted from verbal potency theory that if the word as such was the important element in the naming procedure, then the NN condition should lead to superior retention compared to the PP condition, and should show equal performance to PN and NP, on a recognition task involving equal amounts of both picture and word material. On the basis of the trace modality hypothesis however, rather different results would be expected. Firstly, as recognition is specific to the modality in which an item is encoded, *S*s in the NN condition should show a significantly greater score on the word section of the recognition task relative to the picture section, while the opposite tendency should be present in the results from PP. Secondly, as retention is a function of multiplicity of encoding modalities, then conditions involving the laying down of both visual and verbal traces at presentation (i.e., PN and NP) should show superior performance on the recognition task compared to conditions where visual or verbal traces alone were present (i.e., PP and NN).

METHOD

Subjects. Seventy-two children aged between 9½ and 10½ years and drawn from two local primary schools served as subjects. They were assigned to one of four groups on the basis of age, sex, and error score on the Raven Progressive Matrices Test (Raven, 1956). Children who were considered by their teachers as being unable to read, or whose score on the Matrices Test placed them in the 25th percentile or below, were excluded from the experiment. Each group corresponded to one of the four conditions of presentation. All groups contained nine boys and nine

girls. Age and error score data for the groups of children selected are shown in Table 1. An analysis of variance of this data gave F values <1 for all comparisons.

TABLE 1
MEAN AGE IN MONTHS AND ERROR SCORE ON THE RAVEN MATRICES TEST FOR *Ss* IN THE FOUR PRESENTATION GROUPS

Presentation condition	PP	NN	PN	NP	F
N	18	18	18	18	
Mean age (months)	108.3	108.2	107.4	109.4	<1
Mean errors	8.1	7.8	8.0	7.9	<1

Materials. Two types of material were needed for the presentation task, packs of stimulus cards used by the *E* and response arrays for the *S*. There were two different packs of stimulus cards, one set showing the names of common objects and the other showing pictures of the same objects. Each pack was made up of twelve 5×5-inch cards. A photograph of a different common object was shown on each card of the picture pack. The cards in the name pack showed the names of each of the 12 chosen objects, one to a card. Corresponding to these two packs were the name and picture response arrays. The name array showed all 12 names together on one card, arranged in three rows of four words each. The picture array consisted of a photograph of the 12 objects, similarly grouped. Each array measured 15×10 inches and was encased in a clear perspex wallet for the duration of the experiment.

The 12 items chosen for the presentation task were selected for their familiarity to the *Ss* and the ease with which they could be unambiguously labeled. The items selected were: giraffe, gun, comb, toothbrush, compasses, pen, pipe, car, apple, nail, scissors, and saucer.

In the manufacture of the presentation material an attempt was made to control for stimulus intensity by roughly equating the size of type used for the "name" material to the size of the stimulus objects reproduced in the "picture" material.

For the recognition task, the 12 original stimuli were mixed with 12 new items, which were selected on the basis of their degree of association with the originals. Original stimulus item "comb," for example, could be confused with "brush" in the recognition task, "apple" with "orange," and so on.

Two sets of recognition material were prepared, each set being made up of a name and a picture-recognition array. The picture recognition array showed half the original objects interspersed with half the new,

confusing objects, while the name array showed the remaining old and new items, this time as words. In appearance and format, the recognition arrays resembled the response arrays used for the presentation task. The two sets of material differed in that one was a counterbalanced version of the other, such that items appearing as pictures of objects in one set appeared as names in the second and vice versa.

Procedure. For the *Presentation Task, Ss* were seen individually and in a random order with regard to condition. Experimenter and *S* sat opposite each other across a low table, with the response array face-down in front of the *S*. The *E* told the *S* they were going to play "a matching game"; everytime the *E* showed him a stimulus card, he was to locate and point to the corresponding word or picture on the response array in front of him. Subjects in all groups were cautioned against covert verbalization. The *E* then turned over the *S*'s response array and the presentation task commenced. The *E* picked up the appropriate pack of stimulus cards and showed each card in turn to the child, pausing after every card for the *S* to make the correct response.

Ss in all four groups found little difficulty in mastering their respective response procedures. At the completion of the task, *Ss* were congratulated on their performance and no intimation was given that a test of retention would be carried out at a future date.

The *Recognition Test* was administered four days later. Once again *Ss* were seen individually and in the same order. The *E* reminded the *S* of the "matching game" they had played previously and then presented him with one of the two sets of recognition arrays. The set of recognition material used, and the position of the picture and name components of each set relative to the *S*, were systematically alternated between *Ss*. The *E* then handed the *S* a felt pen and asked him to ring the 12 items he had seen previously. *Ss* were restricted to 12 choices, half to be taken from the name array and half from the picture array. This latter precaution was necessary to ensure that *Ss* in the NN and PP groups did not bias their choices overwhelmingly toward the form of the material with which they were familiar, a tendency shown by some *Ss* in pilot work. When the child had completed the task, irrespective of the level of success, the child was told he or she had "done well' and was sworn to secrecy as to the nature of the recognition task. This precaution against contamination seemed to have been successful as few *Ss* evidenced foreknowledge of the nature of the task in their performance.

RESULTS

The data from the two schools were amalgamated after comparisons to ensure that both samples could be treated as being drawn from a

common population. Mean items correctly recognized on the name and picture sections of the recognition task, arranged according to method of presentation of the original stimuli are shown in Table 2.

TABLE 2

MEAN NUMBER OF CORRECT CHOICES ON THE PICTURE AND NAME SECTIONS OF THE RECOGNITION TASK FOR EACH CONDITION OF PRESENTATION[a]

	Condition			
	NP	PN	PP	NN
Section				
Picture	5.56	5.00	5.00	3.44
Name	5.22	5.06	4.06	4.67
	Tukey comparison of mean total recognition scores			
Mean	10.78	10.06	9.06	8.11
NP	—	ns	$<.01$	$<.01$
PN		—	$<.05$	$<.01$
PP			—	$<.05$
NN				—

[a] Each mean based on 18 cases. The maximum possible score for each section of the recognition test was 6.

A two-way analysis of variance was computed from these data using method of presentation and the form of recognition material as main effects. This produced a highly significant effect for variations in presentation method ($F = 12.32$, $df = 3/128$, $p < .001$). Although the "name" and "picture" recognition material did not differ in their relative effectiveness in facilitating performance overall ($F = .55$, $df = 3/128$, NS), this result concealed wide discrepancies in relative score on the two types of recognition material for different groups, as evidenced by the significant interaction for presentation method with recognition task ($F = 3.21$, $df = 9/128$, $p < .01$).

The significance of differences between total mean scores on the recognition test for the four presentation groups were examined using Tukey's method. The results of this analysis also appear in Table 2. While PN and NP do not differ in terms of total items recognized, both show significantly greater scores than NN and PP. In turn, PP shows significantly superior performance to NN which has the lowest mean score of all.

There is however, considerable support for the views of Wallach and Averbach. The first hypothesis is confirmed. Subjects in the NN and PP conditions score more highly on the type of recognition material with

which they were familiar from the presentation task. On examination the significant interaction between presentation method and recognition task is found to be almost wholly due to differences in relative score on the two parts of the recognition test by *Ss* in the PP and NN conditions. Children in NN scored significantly higher on the name section of the recognition test, relative to the picture section ($t = 2.54$, $df = 34$, $p < .02$), while precisely the reverse occurs in PP where *Ss* score relatively higher on the picture material ($t = 3.30$, $df = 34$, $p < .01$). Support for the second hypothesis is more equivocal. PN and NP are certainly significantly superior to NN and PP in overall score, but much of this superiority seems to represent the failure of these latter two groups to score well in the unfamiliar mode of their respective recognition tasks. The crucial test for a redundancy interpretation is that the mixed modality groups should be superior to the single modality groups even when familiarity with type of material is controlled. Separate Mann Whitney U tests were employed to compare PN and NP with PP on pictures and with NN on names respectively. The results suggested that while the picture and name presentation conditions were not superior to the picture only conditions in the picture material ($Z = 1.04$, NS), they did show such a superiority over the name only conditions in the name material ($Z = 1.78$, $p < .04$, one tail). Neither of the hypotheses derived from verbal potency theory are supported. Subjects in the NN condition not only fail to perform as well as those in PN but also give significantly fewer correct recognitions than PP, where verbal potency theory should have predicted exactly the opposite outcome.

As responses were restricted to 12 per *S*, errors were a simple inverse function of correct identifications, and are consequently not analyzed separately.

DISCUSSION

The results reported here are in agreement with those of Kurtz and Hovland in showing that *Ss* who attach the correct names to objects subsequently recognize them more efficiently than *Ss* who match each of the same objects with its likeness. However, the data from the NN and NP groups suggest that any interpretation of this result in terms of the word as a necessarily superior coder of sense impression is, at least, an oversimplification. On the other hand, the data do provide a measure of support for Wallach and Averbach's trace modality explanation. There is evidence for their first supposition, regarding trace modalities, in that *Ss* experience difficulty in identifying stimuli shown in a different symbolic mode from that used at presentation. But there is only partial support for the second part of their hypothesis, that trace modalities combine

additively to facilitate retention. The failure of PP to show significantly inferior performance to PN and NP on the picture section of the recognition task was also noted by Kurtz and Hovland, who did however find such a difference in terms of errors. The failure of both experiments to find such a difference on a measure of correct identifications, may reflect the opportunities open to *Ss* performing the PP task to covertly verbalize at least some of the names of the objects, thus ensuring sufficient trace redundancy to obscure any significant inferiority for this group, in comparison to PN and NP on picture recognition. There is support for this explanation in the otherwise unexpected finding that PP showed significantly superior retention to NN.

The present results also dispose of two ambiguities arising from the Kurtz and Hovland study. Firstly, overt verbalization by *Ss* in PN is not a necessary prerequisite for the occurrence of the naming effect, though there is some evidence that the effect is marginally enhanced if *Ss* are required to do so (Rosenbaum, 1962). Secondly, the failure of Kurtz and Hovland to differentiate PN from PP on overall recognition performance as opposed to recall does not reflect some basic difference between the two processes as Arnoult (1956) has proposed, but rather the low ceiling level of the recognition task for their *Ss*.

Prompted by the frequent finding that serially presented pictures of objects are retained better than names of the same objects (Kirkpatrick, 1893; Calkins, 1897; Herman, Broussard, and Todd, 1951; Epstein, Rock, and Zuckermann, 1960; Shepard, 1967), Jenkins, Neale, and Deno (1967) have proposed two more sophisticated versions of the verbal potency hypothesis. However, they seem to fit the present data no better than the straightforward version tested here. Their first suggestion is that pictures of objects provoke more extended verbal analysis in the *S* than do names, and that these elaborate descriptions, being richer in association, are then more resistant to extinction than the terse verbal labels occasioned by the sight of the names alone. While this might be plausible in a serial presentation situation, it is more difficult to uphold in the present task where *S*'s response to the stimuli is to a large extent controlled. PP offers least *E* control over *S*'s response yet this produces inferior retention to NP where *S* is given the verbal label he must apply to the object before he ever begins any visual search for it. Their second suggestion, that the different verbal associational characteristics of names and pictures favor the retention of pictures over names also founders on the degree of control exercised over the *S*'s response. Though this theory would correctly predict that NN should show least effective performance, it should surely predict no difference in score between any of the other three groups.

It seems plausible to apply a trace modality type explanation to Jenkins *et al.*'s own data on the retention of serially presented names and pictures. Apart from the usual finding that pictures were retained better than names, they also reported that while *S*s experienced little difficulty in recognizing names previously presented as pictures, they were handicapped in identifying pictures previously shown as names. This situation could arise if *S*s were naming the pictures and only reading the names. These responses would ensure redundant coding for the picture material but only verbal coding for the name material, hence the observed response pattern of the two groups on the recognition test and the superior retention of pictures, as opposed to names of objects.

Despite the suggestive implications of such results, one should hesitate before attempting to extend a trace modality type explanation to all age groups in all situation involving the retention of name and object material. All that is justified at present is to note that the trace modality argument seems to offer a precise and economical explanation for findings on the retention of pictured objects by 10-year-old children. It may well be that above this age, the verbal system plays an increasingly important role in encoding and storing pictured stimuli whereas below 10 years a nonverbal system predominates (e.g., the implications of "eidetic imagery"). It is only to be expected that *S*s will make the most effective use of the coding systems available to them, at different ages. Perhaps research aimed at identifying developmental changes in the relative efficiency of the different labeling procedures described in this paper may shed light on this problem.

REFERENCES

ARNOULT, M. D. Recognition of shapes following paired-associates pretraining. *In* G. Finch and E. Cameron (Eds.) *Symposium on Air Force Human Engineering, Personnel and Training Research.* Washington D. C.: National Academy of Science, National Research Council Publication, 455, 1956, pp. 1–9.

BARTLETT, F. C. *Remembering; an experimental and social study.* London: Cambridge University Press, 1932.

CALKINS, M. W. A study of immediate and delayed recall of the concrete and the visual. *Psychological Review,* 1898, **5,** 451–456.

CARROLL, J. B. *Language and thought.* New Jersey: Prentice-Hall Incorporated, 1964.

EPSTEIN, W., ROCK, I., AND ZUCKERMANN, C. B. Meaning and familiarity in associative learning. *Psychological Monographs,* 1960, **74,** whole no. 491.

ERICKSON, C. W. *Studies in visual perception.* New York: Wiley, 1967.

HERMAN, D. T., BROUSSARD, I. G., AND TODD, H. R. Inter-trial interval and rate of learning serial order picture material. *Journal of General Psychology,* 1951, **45,** 245–254.

JENKIN, A. M. Imagery and learning. *British Journal of Psychology,* 1935, **26,** 149–164.

JENKINS, J. R., NEALE, D. C., AND DENO, S. L. Differential memory for picture and word stimuli. *Journal of Educational Psychology,* 1967, **58,** 303–307.

KATZ, P. A. Effects of labels on children's perception and discrimination learning. *Journal of Experimental Psychology,* 1963, **66,** 423–428.

KIRKPATRICK, A. E. An experimental study of memory. *Psychological Review,* 1893, **1,** 602–609.

KUHLMANN, F. On the analysis of memory consciousness for pictures of familiar objects. *Psychological Review,* 1907, **13,** 316–348.

KURTZ, K. H., AND HOVLAND, C. I. The effect of verbalisation during observation of stimulus objects upon accuracy of recognition and recall. *Journal of Experimental Psychology,* 1953, **45,** 157–164.

RAVEN, J. C. *Coloured Progressive Matrices.* London: H. K. Lewis, 1956.

REESE, H. W. Imagery in paired-associate learning in children. *Journal of Experimental Child Psychology,* 1965, **2,** 290–296.

ROSENBAUM, M. E. Effect of direct and vicarious verbalisation on retention. *Child Development,* 1962, **33,** 103–110.

SHEPARD, R. N. Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6,** 156–163.

SPIKER, C. C. Experiments with children on the hypothesis of acquired distinctiveness and equivalence of cues. *Child Development,* 1956, **27,** 253–263.

WALLACH, H., AND AVERBACH, F. On memory modalites. *American Journal of Psychology,* 1955, **68,** 249–257.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 459–466 (1969)

Noun Concreteness and Verbal Facilitation as Factors in Imaginal Mediation and PA Learning in Children[1]

JOHN C. YUILLE[2] AND SUZANNE PRITCHARD
McGill University

An experiment was conducted to study the effects of noun concreteness and type of verbal connective on speed of discovery of imaginal mediators for Grade-2 and Grade-6 children. A second experiment examined the effects of the same variables on paired-associate (PA) learning. In both experiments concreteness had a facilitatory effect, while the type of connective had no apparent effect. A significant negative correlation was found between mediation latencies and PA learning. These findings were interpreted as evidence for mediating imagery being effective in PA learning, particularly with the older group of children.

In a review of a series of studies, Rohwer (1967) reports that the type of words used to connect pairs of nouns differentially facilitates paired-associate (PA) learning in children: Recall is higher when verbs rather than prepositions are used as connectives, prepositions are more facilitatory than conjunctions; and conjunctions do not differ from using no connective at all. One explanation, of a number of alternatives, that Rohwer (1967) considers, is that the facilitating effect may be mediated by imagery, i.e., nouns linked by a verb may evoke a distinct image which can act as a mediator of the pair. The present study was designed, in part, to test the hypothesis that the sentential facilitation effect is due to mediating imagery.

A series of studies by Paivio and his associates (for a review, see Paivio, in press) bears directly on the question of mediating imagery. Paivio (1965) conducted a PA learning study with adults in which the concreteness-abstractness of stimuli and responses was varied. Recall scores for the four types of pairs decreased in the order: Concrete-

[1] This research was completed while the first author held a Postdoctorate Fellowship from the National Research Council of Canada. The research was supported by a grant from the Social Science Research Council of McGill University.

[2] Now at the University of British Columbia.

concrete (C-C), concrete-abstract (C-A), abstract-concrete (A-C), and abstract-abstract (A-A). This superiority of concrete nouns supported an hypothesis (Paivio, 1963) that the effectiveness of the concreteness attribute is due to mediating imagery. Further support for this imagery hypothesis is provided by the finding that concrete nouns are rated higher than abstract nouns in their capacity to evoke sensory images (Paivio, Yuille, and Madigan, 1967); and that *Ss* report using imaginal mediators most frequently in PA learning when the stimulus or response is concrete (Paivio, Smythe, and Yuille, in press).

Paivio and Yuille (1966) replicated the PA learning results with children. However, the relatively stronger stimulus concreteness effect was less reliable with the younger group than with adults. It was suggested that response learning might be a more important factor in children thus overriding the relative potency of the stimulus. A second purpose of the present study was to examine imaginal mediation in children as a function of stimulus and response concreteness and age.

In order to relate imaginal mediation to the type of verbal links, to age, and to concreteness, it was necessary to find a technique providing "direct" investigation of mediators. Yuille and Paivio (1967) present such a technique. They instructed *Ss* to form mental images linking pairs of words, and, via a key press, measured the latency of mediator discovery. The latencies were found to vary as a function of concreteness, increasing in the order: C-C, C-A, A-C, and A-A. These results suggested that imagery is effective in PA learning since the pattern of latency results is inversely related to the pattern of recall results. This imaginal RT technique appeared well suited to the needs of the present study.

In this investigation, children at two age levels were instructed to find images to link the members of pairs of words. Concreteness of the stimuli and responses was varied as well as the type of verbal link joining the pair (a verb, a conjunction, and no link). On the basis of Rohwer's (1967) interpretation, it was predicted that latencies would be shorter for pairs connected by verbs than by either conjunctions or no link, while the latter would not differ. From another set of studies (e.g., Yuille and Paivio, 1967) it was also expected that concreteness, especially of the stimulus term, would affect latencies, although the effect might be less than obtained with adults (Paivio and Yuille, 1966). Since children under 8 years of age apparently are deficient in verbal mediation ability (Kendler, 1963; Reese, 1962), it was expected that the facilitating effect of the connective form might be found only with concrete material for the younger group.

EXPERIMENT 1

Method

Subjects. The *S*s were 36 Grade-2 (mean age 7.6 years) and 36 Grade-6 (mean age 11.7 years) boys and girls from two schools in a lower-middle class district of Montreal. All *S*s were native English speaking.

PA lists. Two 16 pair lists (lists A and B) were constructed using 16 abstract and 16 concrete nouns. The nouns were those used by Paivio and Yuille (1966). The words had been selected from the 1000 most frequently occurring words listed by Thorndike and Lorge (1944). Paivio and Yuille (1966) had judges rate each noun as concrete or abstract; they also found that the nouns retained their meaningfulness rank across grades. The 32 nouns were randomly paired to form four pairs of each of the stimulus-response combinations of concreteness: Concrete-concrete (C-C), concrete-abstract (C-A), abstract-concrete (A-C), and abstract-abstract (A-A) pairs.[3] The second list was formed by re-pairing the nouns.

Following the method employed by Rohwer (1967) five-word sentences of the form article-noun-verb-article-noun were constructed for each pair of nouns. These constituted the stimuli for the Verb Condition. The Conjunction Condition employed the same five-word strings with a conjunction substituted for the verb in each sentence. In the No Link Condition, the noun pairs appeared alone.

Each "pair" in each condition was printed in black ink on a 5 × 8-inch card, the nouns in upper-case letters, and the other words in lower-case letters.

Apparatus. An electric, clutch-driven clock, with two push-button switches connected in series with the clutch, was used to measure reaction time. The stimulus cards rested on an upright card rack.

Procedure. Each *S* was tested individually in a small room. The instructions informed *S* that the study was concerned with mental images; he was asked to try to formulate and describe some "pictures in his mind." Subsequent to this "practice," *S* was presented three practice pairs and then the 16 cards from one of the three conditions. Twelve *S*s from each grade received one condition. The instructions asked *S* to press a switch as soon as he found a mental picture connecting the words. Upon presentation of a card *E* pressed a switch which started the clock. The clock was stopped by *S*'s key press. After pressing the switch, *S* described

[3] Examples of the four types of pairs: C-C, table-dollar; C-A, cup-spirit; A-C, cost-sugar; A-A, idea-age.

his image. These descriptions were recorded via magnetic tape. If *S* did not find an image within 20 seconds, a latency of 20 seconds was recorded and the next card presented.

The cards were presented in the same random order to all *S*s. Half of the *S*s in each condition received the first list, the other half saw the second set of materials.

Results

During the transcribing of the children's imagery descriptions, it became apparent that not all of them were following the instructions. Some of the children appeared to be responding regardless of whether they used imagery. In such cases, their reaction times would not reflect imaginal latencies, and it was felt that this distortion should be controlled. Three independent judges, naive with respect to this study, were asked to classify each description as "image" or "not image." At the same time, they rated each image in terms of its effectiveness as a mediator of the pair of nouns. The rating was made on a 7-point scale defined by 1—no association, to 7—a close association between the nouns. If two or three judges classified a description as "not image," a latency of 20 seconds was assigned to that pair. At least two of the judges' classifications agreed on 72% of the pairs.

The adjusted mean latencies for each condition appear in Table 1. These results were analyzed in a $2 \times 3 \times 2 \times 2$ analysis of variance. This included two independent factors, grade (two levels) and type of link (three levels), and two repeated measures, stimulus, and response concreteness. A significant grade effect ($F = 23.65$; $df = 1,66$; $p < .01$), indicated that Grade-6 children were faster at discovering images than those in Grade-2. As expected, both stimulus concreteness ($F = 27.03$; $df = 1,66$; $p < .01$) and response concreteness ($F = 26.19$; $df = 1,66$; $p < .01$) facilitated mediator discovery. A significant Stimulus $\times$ Grade interaction ($F = 8.83$; $df = 1,66$; $p < .01$) revealed that the facilitating effect of concrete stimuli was greater for Grade-2 than Grade-6 children. The predicted main effect of type of connector was not obtained ($F = 1.0$).

An analysis of variance of the mean number of images discovered showed exactly the same pattern as the latency data analysis. Similarly, an analysis of the ratings of mediator effectiveness showed the same pattern, except that there was no significant main effect of grade. Since effectiveness, number, and latency of mediators are correlated response measures, the similarity of these results is not unexpected.

The finding that the number of images discovered by each group varied as a function of the experimental variables calls for a re-examination

TABLE 1
MEAN LATENCIES (SECONDS) FOR GRADES 2 AND 6 AS A FUNCTION OF STIMULUS AND RESPONSE CONCRETENESS AND TYPE OF LINK[a]

	Grade 2 response		Grade 6 response	
Stimulus	Concrete	Abstract	Concrete	Abstract
Verb				
Concrete	12.37	15.30	9.14	13.65
Abstract	16.41	17.71	10.61	11.68
Conjunction				
Concrete	12.66	14.78	7.92	8.32
Abstract	14.97	16.89	9.17	11.03
No link				
Concrete	10.56	11.90	9.13	10.33
Abstract	14.11	15.91	10.25	11.55

[a] *SD*s range from 9.49 to 27.36.

of the latency data. Since those pairs for which no image was discovered were given a latency of 20 seconds, the latency of mediator discovery may have varied only because the number of images varied. The mean latency for each type of pair for those pairs which actually evoked an image was calculated for each *S*. The only significant effect involved Grade-6 children discovering images more rapidly than Grade-2 children for all four types of pairs ($t \geq 2.51$; $df \geq 46$; $p < .05$).

These results do not confirm the imagery interpretation of Rohwer's (1967) data. Even for the C-C pairs, which are directly comparable to Rohwer's materials, there is no evidence of the facilitatory effect of type of connective on imaginal mediation. In order to assist the interpretation of these results, a second experiment was conducted in which PA learning data were obtained from similar groups of children, using the same materials.

EXPERIMENT 2

Method

Subjects. *S*s were 18 Grade-2 and 18 Grade-6 pupils, some from one of the schools used in Expt. 1, and some from a third school in the same district.

Materials. The verbal materials and cards of Expt. 1 were used in this study. In addition, two sets of cards, one for each list, were prepared containing the stimulus member of each pair. These cards were used during recall trials.

Procedure. Six *S*s in each grade received one of the three conditions

(Verb, Conjunction, No Link). *Ss* were tested individually. The instructions explained that the task was "to remember which two words belong together, so that when you see the first one, you can tell me the second." After three practice pairs, *S* was given two alternating learning and recall trials with one of the two 16 pair lists. Each card was exposed for 5 seconds during learning. During recall, the card was changed as soon as *S* made a response. *E* informed *S* whether the response was correct. The material was read aloud by *E* during both learning and recall. The order of presentation was randomly varied.

Results

The mean total correct recall scores over two trials for each condition appear in Table 2. A $2 \times 3 \times 2 \times 2$ analysis of variance was performed on these data. Recall was higher for Grade-6 than Grade-2 children ($F = 5.80$; $df = 1,30$; $p < .05$). As expected from previous results, both stimulus concreteness ($F = 48.09$; $df = 1,30$; $p < .01$) and response concreteness ($F = 18.88$; $df = 1,30$; $p < .01$) facilitated recall performance. The interaction of stimulus and response variables ($F = 4.60$; $df = 1,60$; $p < .05$) was due to concrete stimuli having a greater facilitative effect when the response was concrete. A significant Grade $\times$ Response interaction ($F = 9.87$; $df = 1,30$; $p < .01$) revealed that the facilitating effect of response concreteness was largely located in the Grade-2 conditions. The expected facilitating effect of the Verb Condition was not obtained ($F = 1.0$). However, a Stimulus $\times$ Condition interaction ($F = 5.32$; $df = 2,30$; $p < .05$) revealed that variation in stimulus

TABLE 2

MEAN TOTAL RECALL FOR TWO TRIALS AS A FUNCTION OF STIMULUS AND RESPONSE CONCRETENESS AND TYPE OF LINK[a]

	Grade 2 response		Grade 6 response	
Stimulus	Concrete	Abstract	Concrete	Abstract
Verb				
Concrete	3.83	2.33	4.83	3.83
Abstract	2.50	1.50	2.83	2.67
Conjunction				
Concrete	5.00	2.17	4.17	4.00
Abstract	2.00	1.50	2.33	2.17
No link				
Concrete	3.50	1.50	4.33	4.33
Abstract	3.33	1.83	3.33	3.33

[a] *SD*s range from 0.58 to 2.22.

concreteness had a greater effect with conjunction links, and least effect with no links.

DISCUSSION

Children, of the type tested in this study, apparently discover imaginal mediators for noun pairs with equal rapidity regardless of the presentation context. This suggests that imagery may not be the mediating process underlying Rohwer's sentence effect. However, the failure to replicate Rohwer's PA learning results with a similar group of children makes conclusions about the sentence effect difficult. Perhaps the reason for not replicating the results is due to differences in materials. The low recall scores obtained in the present study suggest that the pairs were difficult to learn. This may have affected the usefulness of connectives. In addition, population and procedural differences were involved. For example, Rohwer presented his materials verbally while a combined visual-verbal method was employed in this study. Further research is necessary before the reasons for this difference in results can be assessed.

The strong effect of concreteness on imaginal mediation latency in children extends the findings previously reported with adults (Yuille and Paivio, 1966). The stronger stimulus effect obtained with Grade-2 rather than the Grade-6 group is unexpected. Perhaps the younger children attended only to the first noun of each pair, while the Grade-6 children, more experienced with words, can take advantage of concreteness in either location. The fact that rated mediator efficiency was more affected by stimulus concreteness in the Grade-2 group supports this interpretation.

The PA learning results are similar to those obtained with adults. As with previous research with children (Paivio and Yuille, 1966) the stronger stimulus effect was not clear. The fact that the response variable had a relatively greater effect with the younger group may be due to response learning (cf., Underwood and Schulz, 1960) rather than mediation.

One implication of this study is that imaginal mediation ability increases from Grade 2 to Grade 6. Grade-6 children form more images, and form them faster than do the Grade-2's. It also seems that the older children use the mediators to better advantage. Some evidence for the fact that imagery was used by the children is revealed by the correlations between mediation and learning data. The Pearson product moment r between mean adjusted latency (Table 1) and recall for each pair of nouns, collapsed over type of link, is $-.59$ ($p < .01$), ($-.28$ for Grade 2, $-.48$ for Grade 6). Also the number of *Ss* finding an image for a pair correlates .36 ($p < .05$) with the recall for that pair. This

evidence, although inferential, supports the hypothesis that imagery can serve as an effective mediator of PA learning.

REFERENCES

KENDLER, T. S. Development of mediating responses in children. *Child Development,* 1963, **28,** 33–48.

PAIVIO, A. Learning of adjective-noun paired-associates as a function of adjective-noun word order and noun abstractness. *Canadian Journal of Psychology,* 1963, **17,** 370–379.

PAIVIO, A. Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4,** 32–38.

PAIVIO, A. Meaning, mediation, and memory. *Psychological Review,* (in press).

PAIVIO, A., SMYTHE, P. C., AND YUILLE, J. C. Imagery versus meaningfulness of nouns in paired-associate learning. *Canadian Journal of Psychology,* (in press).

PAIVIO, A., AND YUILLE, J. C. Word abstractness and meaningfulness and paired-associate learning in children. *Journal of Experimental Child Psychology,* 1966, **4,** 81–89.

PAIVIO, A., AND YUILLE, J. C. Changes in associative strategies and paired-associate learning over trials as a function of word imagery and type of learning set. *Journal of Experimental Psychology,* (in press).

PAIVIO, A., YUILLE, J. C., AND MADIGAN, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology,* 1968, **76,** No. 1, pt. 2 (supplement).

REESE, H. W. Verbal mediation as a function of age level. *Psychological Bulletin,* 1962, **59,** 502–509.

ROHWER, W. D., JR. Social class differences in the role of linguistic structures in paired-associate learning: elaboration and learning proficiency. U. S. Office of Education, Bureau of Research Project No. 5-0605, 1967.

THORNDIKE, E. L., AND LORGE, L. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J., AND SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

YUILLE, J. C., AND PAIVIO, A. Latency of imaginal and verbal mediators as a function of stimulus and response concreteness-imagery. *Journal of Experimental Psychology,* 1967, **75,** 540–544.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 467–478 (1969)

An Attempt to Shape Bidimensional Attention in 24-Month-Old Infants[1]

JOHN S. WATSON
University of California, Berkeley

AND

GARY DANIELSON
Florida State University, Tallahasse

Twenty-four *Ss* within 2 weeks of 24 months of age were presented a shaping procedure designed to lead to bidimensional sorting behavior. The shaping sequence involved three steps. Step I required spatial separation of two objects and 22 *Ss* reached criterial performance within the 25 trials allotted to this step. Step II required unidimensional sorting of two objects by color for half the *Ss* and by form for the other half. Ten *Ss* reached criterial performance within the 35 trials allotted to the second step. Step III required sorting three objects by both color and form. Four *Ss* succeeded on the third step at a level sufficient to reject statistically an existing hypothesis that children are limited to unidimensional attention prior to about 7 years of age. Success on Step III was significantly related to success on Step II. Errors on Step III were examined to further assess the unidimensional hypothesis and some support was found. Overall, however, data from this and other studies are viewed as more consistent with an alternative hypothesis that unidimensional and multidimensional attention are in part situationally dependent on the absence or presence of reinforcement for object discrimination.

Within a wide variety of cognitive developmental topics, much is made of an assumption that young children tend to center attention on single dimensions or attributes of stimulus objects. The ability to simultaneously consider two or more stimulus dimensions is presumably absent until about the age of 7 years—at least this appears to be the position presented within the Piagetian literature (Wallace, 1965; p. 80).

Advancement beyond unidimensional attention is viewed as necessary for a variety of significant cognitive functions. Within the writings of

[1]The research reported here was supported in part by OEO Head Start Subcontract No. 1410 with Michigan State University Head Start Evaluation and Research Center. Data collection for the study was carried out while the senior author was on the faculty of and the junior author was a student at the Merrill-Palmer Institute in Detroit.

Piaget, the numerous "conservations" (e.g., of quantity, of area, of social role), nonegocentric spatial and social conceptualization, and multiple-classification behavior are each assumed to be prohibited by the young child's limitation to unidimensional attention. A similar proposal has recently been made outside the Piagetian domain to explain the errors of 6-year-olds in a study of the development of conjunctive and disjunctive rule formation (King, 1966).

Direct support for the contention that young children are limited to unidimensional attention comes primarily from two sources of data. These are the behavior of young children in object-sorting tasks and in Piagetian conservation tasks. An example from conservation tasks is Piaget's observation that when a young child watches a piece of clay being elongated into the shape of a sausage, the child will consider the change in shape to involve a change in amount of clay. If he centers his attention on length, he will think the clay is increasing in amount. If he attends to the fact that the sausage becomes thinner, he will think it is decreasing in amount. "In both cases, the child is unaware of conservation, and refers to only one dimension, either one or the other, not both at the same time" (Piaget, 1967, p. 155).

The behavior of young children in object-sorting tasks has also provided direct support of a limitation to unidimensional attention in children below the age of about 7 years. Children from about 2 to 7 years of age appear quite able to sort objects on the basis of a single attribute, but they "cannot employ two attributes of the same object, that is, break up a group of apples along the multiple dimensions of big red apples and small green apples" (Sigel, 1964, p. 218). A notable exception to this general finding occurred in a study by Colby and Robertson (1942) which was designed to assess the relative frequency of color and form serving as a consistent basis for sorting by children ranging from 3½ to 9½ years old. A few children younger than 5 years of age were judged capable of sorting on basis of color, and size simultaneously.

Although Colby and Robertson did not provide an estimate of the chance probability of their exceptional cases, if one assumes they are valid cases of multidimensional attention then at least two different theoretical interpretations can be proposed to account for them. On the one hand, these cases can be viewed as an "illustration of the need to be sensitive to the individual differences in the stages in which various phenomena appear, rather than being overly dependent on age levels" (Sigel, 1964). This position accepts the cases as early arrivals in the developmental stage which embodies the ability to attend to multiple dimensions (e.g., the stage of "concrete operations" in Piaget's theory).

An alternative interpretation of Colby and Robertson's cases is that they illustrate a need to distinguish between what children *do* and what they are *capable of doing*, a distinction which Shantz, 1966, has pointed out is often lacking in much of Piaget's work. If some children are able to attend to multiple dimensions before 5 years of age, then the question arises as to whether the children judged unable are truly unable or if perhaps the standard assessment procedures have simply failed to elicit an existing capacity for multidimensional attention. In this regard, it is notable that the standard sorting and conservation tasks are noninstrumental in form; that is, they do not provide reward for correct responses.

Thus, existing data at best support the contention that children younger than 7 years of age normally employ unidimensional attention in noninstrumental discrimination tasks. Whether this tendency toward unidimensional attention is also normally dominant in instrumental discrimination tasks is not yet empirically supported. Even if it were to be found that unidimensional attention is a normally dominant response tendency for both instrumental and noninstrumental discrimination tasks an important question would remain as to whether multidimensional attention is beyond the functional capacity of the young child or if it is rather a response which can be learned under appropriate conditions.

The present study attempted to shape bidimensional attention in 24-month-old *S*s within a sorting task involving objects which vary on the basis of two levels of the dimension of color and two levels of the dimension of form. The study had two principal objectives. One was to assess the extent of success of this technique at this age. The other objective was to assess the extent to which unidimensional attention accounts for errors in this instrumental discrimination task.

METHOD

Subjects. The *S*s were children within 2 weeks of 24 months of age. They were solicited by mail from published lists of official birth announcements for the Detroit area. Potential *S*s were selected from these lists on the basis of sex and age until 6 boys and 6 girls within each of two experimental groups met certain performance criteria (i.e., completed Step II of the shaping procedure described below) for entrance into the participating sample of the study. Two boys and ten girls failed to meet these performance criteria. Thus, sample construction required successful solicitation of 36 *S*s and these represent the total positive responses to approximately 300 soliciting letters. All *S*s except one were Caucasian. The educational level of the fathers of the 36 *S*s ranged from 10 to 20 years of formal schooling with a mean of 15.5 years.

Stimulus objects. The various sorting tasks involved use of three

plastic blocks each of which was ½-inch thick with a surface area of approximately 2½ square inches. The blocks differed from one another on the basis of the specific values (black vs. red and circle vs. triangle) of two binary dimensions (color and form). The specific combinations employed were black-circle, black-triangle, and red-circle.

Apparatus. Sorting behavior was elicited and rewarded with the use of the sorting box shown in Fig. 1. The box has two slots 12 inches

FIG. 1. The sorting box.

apart on the left-right axis near the top of its slanting front surface. Blocks placed in these slots are retained within the box until released by an electric switch. When released, the blocks enter a "return basket" positioned mid-line at the front base of the box. The head and torso of a clown are painted on the slanting front surface. The clown's eyes and nose are 12-volt panel lights which can be lit in unison. Within the box are an electric train whistle and a Trix cereal dispenser which can deliver a single piece of cereal into a "reward basket" situated next

to the object return basket. The lights, train whistle, and cereal served as reinforcement as described below.

PROCEDURE

The *S*s were tested individually in a laboratory room at the Merrill-Palmer Institute. The mother of each *S* accompanied the child into the room and was seated approximately 10 ft from the sorting box. As soon as *E* judged that *S* had become reasonably adapted to the situation, *S* was given two blocks. The specific blocks used differed for two experimental groups. The blocks were the black-circle and red-circle for Color Group *S*s and black-circle and black-triangle for Form Group *S*s.

Step I of shaping. After *S* examined the blocks, *E* asked for them back saying "Let me show you what I can do with these." *E* then slowly placed one block in the right-hand slot and then one in the left-hand slot. At this point, *E* began blinking the lights on the box by means of a remote control switch held behind his back (the switch provided independent control of the three reinforcements and the object return mechanism). The lights were blinked for 5 seconds, then the object return mechanism was activated. *E* acted elated and asked *S* "Can you do that, can you make the lights go?"

As *S* began putting the blocks in the box, *E* seated himself about 8 feet to one side and began recording *S*'s pattern of object sorting. During Step I of shaping, reinforcement was contingent on spatially separating the two blocks without regard to which block was placed in which slot. If *S* made the error of placing both blocks in one slot, no reinforcement occurred during a period of 5 seconds prior to return of the blocks. Step I continued until *S* was correct on six consecutive trials following the fifth trial or until completion of 25 trials.

Step II of shaping. Step II began directly upon completion of Step I and involved solely a change in the sorting pattern required. For Color Group *S*s, reinforcement was now contingent on a specific pattern of separation; e.g., red-left and black-right. The same was true for Form Group *S*s; e.g., circle-left and triangle-right. These specific position requirements were reversed for half the *S*s in each group. Error was therefore an instance of either not separating the two blocks or reversing the specified pattern of separation. Step II continued until *S* was correct on seven consecutive trials following the fifth trial or until completion of 35 trials.

Step III of shaping. At the start of Step III, *E* showed *S* the third block (i.e., the red-circle for Form Group *S*s or the black-triangle for Color Group *S*s), then demonstrated slowly the correct sorting pattern.

For all *S*s, the correct pattern was specified as placing the black-circle in the slot which had been correct in Step II and placing the red-circle and black-triangle in the other slot and thus correct placement of the black circle required attention to both its color and form. Four types of error were now possible: nonseparation (all three blocks in one slot), reversal (correct separation but wrong slots), form separation (circles in one slot and triangle in the other), and color separation (black blocks in one slot and red block in the other). Criterion was set at eight consecutive correct trials, but with the exception of one *S* Step III continued until *S* stopped playing. This *S* was the first to reach criterion and was stopped in accordance with an initial decision to stop at criterion. However, it soon became clear that criterial performance was infrequent, so this policy was dropped to insure maximal sensitivity for assessments of chance occurrences.

The specific shaping sequence employed in this study was chosen for several reasons. First of all, a *S* could theoretically move through all three steps without facing a greater than .5 chance of making an error. Step I provides four logical alternatives of which 2 are correct. Step II provides the same alternatives, but a *S* mastering Step I should be limited to the two separation patterns of which one is now correct. With mastery of Step II, a *S* should presumably face the task of learning in which of the two slots the new block must be placed. Thus, while there are eight possible ways of sorting the three blocks in Step III, a fully shaped *S* would only face a .5 chance of committing an error.

The shaping sequence employed provided the possibility of introducing either a color or form sorting task at Step II in the sequence. The two types of unidimensional discrimination therefore could be compared with respect to their relative difficulty and relative transfer value for performance on Step III.

Reinforcement. All *S*s began with the lights serving as reinforcement for correct trials. In an attempt to obtain a maximal number of sorting trials, the nature of the reinforcement was changed as soon as *S* met an adaptation criterion of one intertrial delay greater than 5 minutes, or two delays exceeding 2 minutes each, or three delays exceeding 1 minute each. When a *S* met this criterion for the first time, a correct sorting trial was demonstrated and reinforcement was changed from lights to lights plus train whistle. The second time the adaptation criterion was met, a correct trial was demonstrated and reinforcement was changed from lights plus whistle to lights plus whistle plus Trix cereal. When *S* met the adaptation criterion a third time, the session ended.

During all delays exceeding 15 seconds, *E* activated the object return

mechanism which made a "click." This was done every 15 seconds in conjunction with a verbal request by *E* for *S* to continue, until *S* either began sorting again or until the adaptation criterion was met.

RESULTS

Step I. Of the 24 *S*s in the participating sample of this study, 22 reached criterion on Step I within the 25 trials allotted. Criterion was reached without error by 18 *S*s and with but one error by the remaining 4 *S*s. The two failing *S*s were a boy and a girl in the Color Group.

Step II. Within the 35 trials allotted to Step II, 10 of the 24 *S*s reached criterion. As was the case in Step I, those *S*s who reached criterion tended to do so with very few errors (mean = 3.8). Regarding the distribution of *S*s reaching criterion in Step II, it is notable that the ten succeeding *S*s were evenly distributed by sex, but the distribution by method was seven on form and three on color. Thus, the data show a trend suggesting a lesser difficulty of the form sorting task.

Step III. Within the unlimited trials allotted to Step III, four *S*s reached criterion. The *S*s were two Form Group boys, one Color Group boy, and one Color Group girl. Errors to criterion were 43, 1, 1, and 15 respectively. The infrequency of criterial performance on the bidimensional task raises the question of whether these successes are simply products of chance. That is to say, is the performance of these four *S*s sufficient to reject the universal negative hypothesis that no 24-month-old can succeed in Step III? The data support rejection of the universal negative hypothesis on the basis of either of two analyses. On the one hand, since *S*s were permitted to continue sorting indefinitely this allowed one *S* to accumulate one set of 18 consecutive correct trials. Assuming that chance performance above unidimensional sorting has a per trial probability of .5 during Step III, and considering the number of opportunities for one or more sets of 18 or more consecutively correct trials within the responses of all 24 *S*s, the probability of the chance occurrence of the one set of 18 consecutively correct was calculated as less than .003.

Moreover, if performance on the final 14 trials in Step III is considered, a binomial test of the frequency of correct trials exceeded the .006 level for each of the 4 *S*s who reached criterion. Thirteen *S*s continued for at least 14 trials on Step III. It is extremely unlikely that of these 13 *S*s 4 would exceed the .006 level by chance.

Accepting the four *S*s as valid cases of bidimensional sorting, an important question arises as to why these four succeeded while the remaining 20 *S*s did not. One variable of obvious importance is the number of sorting trials which were produced during Step III. The variation

across *Ss* ranged from 0 to 94 trials. The fewest trials to criterion for any *S* was 14 trials. It would seem reasonable, therefore, to eliminate from further analyses the 11 *Ss* who did not accrue at least 14 trials in Step III. The remaining 13 *Ss* thus provide a basis for examining variables potentially related to success within the performance of *Ss* who persevered for at least a minimally sufficient number of trials.

With these 13 *Ss*, comparisons were made between the 4 successes and the 9 failures regarding education of father, sex of *S*, color vs. form task on Step II, and verbal fluency (as based on frequency, structure, and articulation of speech during the testing session). No appreciable trends were evident for any of these variables, though sensitivity, of course, is limited by the small *N* available. However, when success vs. failure on Step II was considered as shown in Table 1, a Fisher Exact Probability Test was significant at the .05 level.

TABLE 1

CONTINGENCY ANALYSIS OF RELATIONSHIP BETWEEN CRITERION PERFORMANCE ON STEP II AND STEP III FOR *Ss* CONTINUING FOR 14 OR MORE TRIALS ON STEP III

		Criterion met on Step II		
		Yes	No	Total
Criterion met on Step III	Yes	4	0	4
	No	3	6	9
Total		7	6	13

Error analysis. Errors during Step III were analyzed to see if they occurred in a manner consistent with the assumption that young children center attention on a single dimension. This study was designed to provide two assessments of the unidimensional hypothesis. First, if a *S*'s errors on Step III were produced by unidimensional attention, then his pattern of errors should reflect this by a dominance of color or a dominance of form error types. Secondly, a tendency toward unidimensional attention could be expected to be strengthened by reinforcement during the unidimensional sorting task of Step II. Were this so, Color Group *Ss* should tend toward color dominance in Step III and Form Group *Ss* should tend toward form dominance—at least this should be expected at the beginning of Step III for those *Ss* reaching criterion on Step II.

The occurrence of unidimensional dominance during Step III was assessed by examining the degree of disproportion in the frequencies of color and form errors for each *S*.

In a previous study by King, 1966, *Ss* had to identify positive and negative instances of conjunctive rules (e.g., red and square) and dis-

junctive rules (e.g., red and/or square) with respect to variations in two visual dimensions of test stimuli. There were four possible cases of test stimuli which were distinguished by the presence (P) or absence (A) of the relevant stimulus attributes, e.g., red-square (PP), red-nonsquare (PA), nonred-square (AP), and nonred-nonsquare (AA). The analysis of unidimensional dominance was limited to comparing the frequency of errors on the PA and AP cases. In the conjunctive problem, PA and AP are negative instances; but a *S* attending to only one dimension (e.g., color) would tend to identify erroneously one of these but not both (e.g., PA but not AP) as a positive instance of the rule. In the disjunctive problem, PA and AP are positive instances, but a *S* attending to only one dimension (e.g., form) would tend to identify erroneously one of these but not both (e.g., PA but not AP) as a negative instance. In both conjunctive and disjunctive problems, then, attending to only one dimension should be reflected in an unequal frequency of PA and AP errors. A *S* attending to both dimensions, on the other hand, should tend to make an equal number of PA and AP errors.

Following analogous reasoning, the color and form errors on Step III of the present study were examined to determine any disproportion in frequency within the individual response records of *S*s. Unidimensional attention would be expected to lead to an unequal frequency of these two error types, but they would be expected to be about equally frequent for a *S* who attended to both the color and form dimensions.

The criterion of unidimensional dominance used by King was not employed here, however, since that criterion incorporates a notable statistical bias. King limited his analysis to *S*s who made at least two errors on the PA and/or AP cases and unidimensional dominance was claimed for *S*s displaying more errors on one case than the other when the ratio of error frequency was 2:1 or greater. Yet *S*s making two, three, or four such errors can be expected to reach this criterion of dominance by chance at the probability rates of .5, 1.0, and .6 respectively. For example, a *S* making three errors has a chance probability of 1.0 of meeting the criterion because there are only eight possible ways of making three such errors (i.e., PA-PA-PA, PA-PA-AP, PA-AP-PA, AP-PA-PA, PA-AP-AP, AP-PA-AP, AP-AP-PA, or AP-AP-AP) and each case meets the criterion of a ratio of 2:1 or greater.[2]

[2] With this criterion of unidimensional dominance, King (1966) found a greater proportion of 6-year-olds than older *S*s showing dominance. Yet without some indication that the groups did not also differ in their distributions of AP and PA error frequencies, King's finding is difficult to interpret. For example, if the case of three or four errors occurred more frequently among 6-year-olds than older *S*s and the case of two errors occurred more frequently among older *S*s than among 6-year-olds, the dominance classification would simply be more probable by chance for 6-year-olds.

In the present study, the analysis of unidimensional dominance was limited to *Ss* making four or more color and/or form errors, and the probability of error ratios as disparate or greater than that observed was calculated for each *S*. Thus, for example, if a *S* made four errors of which three were color and one was form, the probability for a disparity as great or greater than the observed 3:1 ratio would be calculated by determining that of the 16 possible error distributions of 4 errors two are 4:0 ratios and eight are 3:1. Thus, 10 of 16 or .625 of the possible outcomes would be equal to or greater than the observed outcome. The observed case would then be considered as a disparity having the probability of .625 of being reached by chance. Were the observed outcome all four color errors, the chance probability would be .125 since there are only 2 of the 16 possible outcomes that reach the 4:0 ratio of disparity.

Nine of the 13 *Ss* who persevered for 14 or more trials in Step III committed at least four errors of the color or form types. The probabilities for the observed error frequency disparities ranged from 1.0 to .02 across the 9 *Ss*. Six of the nine probabilities were less than .2. A conservative estimate of the liklihood that six of nine *Ss* would show this extent of unidimensional dominance by chance is $p < .01$, as estimated by references to the Poisson distribution.

For five of the six *Ss* showing dominance at $p < .2$, color was the predominant error type. Regarding the question of whether dominance in Step III would be related to the dimension involved in Step II, no transfer effect was observed in that three cases of dominance favored the dimension involved in Step II and three cases did not.

An analysis of only the first four form or color errors was made to check whether unidimensional dominance and transfer effects might be exhibited most strongly within the initial trials of Step III. Contrary to the expected, unidimensional dominance appeared less evident within the initial four errors than in later errors. Of the nine *Ss* who made at least four relevant errors, only two *Ss* showed a uniform consistency of error type within their first four errors. Moreover, of the six *Ss* who eventually developed a notable dominance (i.e., p of ratio $<.2$), four *Ss* had lower ratios initially than later while only one *S* had a higher ratio initally. As for the queston of transfer effects from Step II within initial Step III errors, none was apparent.

DISCUSSION

The results of the present study provide one rather clear and important implication. By the age of 24 months, some children are quite capable of performing a bidimensional sorting task—at least this seems

clear within the stimulus and reward context employed here. While 4 successes out of a starting sample of 36 *S*s may at first appear a modest ratio of capability, it is notable that when reference is limited to *S*s who both mastered the unidimensional task of Step II and continued for at least 14 trials in the bidimensional task, the ratio rises to four successes out of seven *S*s.

Previous discussions of the young child's presumed inability to perform tasks requiring multidimensional attention have viewed this inability as the consequence of the child's *general* limitation to unidimensional attention. Since that proposal has arisen primarily from observations of young children in noninstrumental tasks, error analyses were undertaken to see if the proposal appeared applicable to performance on the instrumental task of this study. Although a significant degree of unidimensional dominance was found in the color and form errors produced during the bidimensional task of Step III, two aspects of the dominance patterns which occurred seem inconsistent with the general unidimensional hypothesis.

On the one hand, the specific direction of dominance was apparently unaffected by the type of unidimensional training received by a *S* during Step II. This might be viewed as reflecting an immutability of an original dominance tendency which *S* brought into the experimental session. Yet this seems unlikely since even *S*s who rapidly mastered the unidimensional task on Step II and thus could be presumed to be displaying their original dominance did not show any consistency of dominance of that dimension within Step III.

Another inconsistency with the unidimensional hypothesis is the finding that unidimensional dominance tended to be less clear within the first four relevant errors of Step III than during later errors. Why would *S*s show less tendency toward unidimensional attention immediately after Step II trials which provided reward for unidimensional attention than after a set of Step III trials in which only bidimensional attention was rewarded? One possible explanation would be that multidimensional attention in the young child is partially under the control of a cue aspect of reinforcement which marks an occasion in which multidimensional attention has been rewarded more consistently than unidimensional attention. If so, then as trials continued in Step III with little or no reinforcement occurring, multidimensional attention would be expected to give way to unidimensional attention. This follows if one assumes unidimensional attention is less complex a response than multidimensional attention and further assumes some validity for the principle of least effort.

This hypothesis of the cue value of reinforcement would also explain

the fact that *Ss* reaching criterion on Step II were significantly more likely to reach criterion on Step III. These *Ss* would enter Step III with a greater tendency toward multidimensional attention due to the immediately preceding high rate of reinforcement received in Step II. Surely simpler explanations are possible; e.g., proposals of ability selection in Step II or shaping consistency from Step II to III. However, it is interesting to note that the hypothesis at hand also fits results of a study by Eimas (1965) in which kindergarten children were found to use compound cues in a discrimination task in which reinforcement could be obtained by attention to single component cues alone.

Upon reflection, it seems likely that many of the learned adaptations which youig children acquire in language and social behavior require attention to both specific cues and their stimulus contexts; e.g., discriminations between intonational contrasts or syntactical contexts of certain words, between love-pats and soft-spanks, or between "warning" from parents in public versus at home. In light of the probable number of such multidimensional discriminations required of young children, it would hardly be surprising if multidimensional attenton were to become eventually the model rather than the rare initial response to instrumental discrimination situations.

It is not clear whether the limited number of *Ss* showing bidimensional attention in the present study is a sign that this response is yet rare at 24 months of age or whether the limited number is an artifact of a failure to obtain and maintain sufficient involvement in the task. However, the performance of the 4 *Ss* who did master the bidimensional task is certainly sufficient to conclude that bidimensional attention is an available response for some children long before their seventh year.

REFERENCES

Colby, M. G., and Robertson, J. B. Genetic studies in abstraction. *Journal of Comparative Psychology,* 1942, 33, 385–401.

Eimas, P. D. Stimulus compounding in the discrimination learning of kindergarten children. *Journal of Experimental Child Psychology,* 1965, 2, 178–185.

King, W. L. Learning and utilization of conjunctive and disjunctive classification rules: A developmental study. *Journal of Experimental Child Psychology,* 1966, 4, 217–231.

Piaget, J. *Six psychological studies.* New York: Random House, 1967.

Shantz, Carolyn U. A developmental study of Piaget's theory of logical multiplication. Unpublished doctoral dissertation, Purdue University, 1966.

Sigel, I. E. The attainment of concepts. *In* M. L. Hoffman and Lois W. Hoffman (Eds.), *Review of child development research.* Vol. 1. New York: Russell Sage Foundation, 1964. Pp. 209–248.

Wallace, J. G. *Concept growth and the education of the child.* Hove, Sussex: National Foundation for Educational Research in England and Wales, 1965.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 479–484 (1969)

The Effect of Verbalization on Discrimination Learning and Transfer in Nursery School Children[1]

L. R. GOULET
West Virginia University

Nursery school children were compared on two-choice discrimination and transfer lists under instructions to verbalize (or not to verbalize) the stimuli before making a choice. The transfer paradigm involved retaining the "old" wrong (List 1) stimulus items and pairing them with "new" right items on List 2, relative to mixed-list control pairs consisting of pairs of new wrong and new right items on List 2. Strong positive transfer was obtained on List 2 under both verbalization and non-verbalization treatments. In addition, there was a strong trend toward facilitative effects of verbalization on both List-1 learning and List-1 relearning after interpolated List-2 practice.

A consideration of the available theoretical propositions relevant to verbal-discrimination learning highlights an apparent discontinuity between young children and adults in terms of the variables affecting discrimination learning. For example, overt verbalization of task items is generally considered to have facilitative effects on discrimination learning of children, whereas deleterious effects of this variable are found in verbal-discrimination (VD) tasks when college students (Kausler and Sardello, 1967; Sardello and Kausler, 1967) or children as young as those in fifth-grade (Goulet and Hoyer, 1969) serve as *Ss*.

The frequency theory of VD learning (Ekstrand, Wallace, and Underwood, 1966) suggests that verbalization of right (R) and wrong (W) items should have detrimental effects in acquisition. The theory is based on the assumption that VD learning is in part a function of the frequency with which the R and W items have been overtly or covertly verbalized during acquisition. The exposure of the items adds a "frequency unit" to both R and W items with the R item adding additional frequency units during the feedback exposure and if rehearsal of the R item occurs. The deleterious effects of verbalization are assumed to occur because frequency units are added to both R and W items rather than

[1]This research was supported in part by NIMH grant MH-13515 to the author. Appreciation is expressed to James Joseph who ran the *Ss* and to Mrs. April V. Bevers, Acting Director of the West Virginia University Laboratory Nursery School for her helpful assistance in the conduct of this research.

to the R item alone and because the time for rehearsal of the correct (R) items is minimized. The present study was concerned with the effects of verbalization of R and W items in the discrimination learning of nursery school children on both an initial and a transfer list.

There has been extremely little research related to the mechanisms of transfer in discrimination learning of young children. The reasons for the paucity of research are not readily apparent since the use of transfer designs can assist in the determination of the basic *learning* and/or associative processes in children. For example, Cross and Vaughter (1966) have provided suggestive evidence that very young children (e.g., CA = 4 years) respond in a two-choice object discrimination task by *avoiding* the non-reinforced (incorrect) stimulus rather than *approaching* the reinforced (correct) stimulus. They used a modified discrimination-transfer task where *S*s were given two training trials involving positive information (forced-choice of the correct stimulus in the absence of the incorrect stimulus), or two negative information trials (forced choice of the incorrect stimulus in the absence of the correct stimulus). Performance on a third (transfer-test) trial where both the incorrect and correct stimulus were available, was better following negative information trials than positive information trials.

A test of this proposition may also be made using a conventional transfer design and procedure; i.e., *S*s may learn two discrimination lists, each consisting of a series of two-choice pairs involving wrong (W) and right (R) items. The first-list pairs may be designated W_1-R_1. A second list may be constructed with "new" R items while still retaining the "old" (List 1) W items; i.e., the W_1-R_1, W_1-R_2 paradigm. If the assumption by Cross and Vaughter is correct, strong positive transfer should be manifest on the second list relative to control (W_2-R_2) pairs where no specific interlist transfer is possible.

Frequency theory does provide additional predictions regarding acquisition of transfer lists. For example, in the W_1-R_1, W_1-R_2 paradigm, *S*s may respond on the basis of frequency cues on List 2 by choosing the item in each pair which has the least frequency units (Rule 2, Ekstrand *et al.*). However, since frequency units accrue more rapidly for R items than W items, discriminative cues based on frequency become unreliable during the later stages of List-2 practice. Underwood, Jesse, and Ekstrand (1964) provided direct support for a frequency analysis of VD learning; i.e., strong positive transfer was obtained for W_1-R_2 pairs on List 2 in the early practice trials, followed by inhibitory (negative transfer) effects on later trials. The negative transfer effects were attributed to the unreliability of frequency cues in the later stages of List-2 practice. Kausler and Dean (1968) provided a basic replication of the Underwood *et al.* study and found that the inhibitory effects were

present throughout List-2 practice if *Ss* were not informed of the exact relation between List 1 and List 2 prior to practice on the second list. Both Underwood *et al.*, and Kausler and Dean used college *Ss*. Presumably, the initial positive transfer found by Underwood *et al.* was due to *Ss* "set" to respond by using Rule 2 rather than using Rule 1 (choosing the item with the greatest frequency units) because of having full information (through instructions) of the List 1–List 2 relations. The present study investigated transfer in nursery school children using the W_1-R_1, W_1-R_2 paradigm.

METHOD

Design and Subjects

The basic design was a 2×2 "mixed" factorial with verbalization (pronunciation of R and W items or no pronunciation) as a between-subject variable and type of pair, W_1-R_2 (E) or W_2-R_2 (C), as a mixed-list factor. The C pairs (new right *and* wrong items on List 2) were included as a baseline against which to assess the effects of specific transfer (Underwood *et al.*) for the E pairs. As a control for possible confounding of pair difficulty in transfer, two sets of lists were constructed such that the second-list E pairs in one set were the C pairs in the second set and vice versa.

The *Ss* were 20 children enrolled in the West Virginia University nursery school who had not previously participated as subjects. The mean age of the children was 4.25. No child was older than 4 years 11 months nor younger than 3 years 5 months. There were 10 boys and 10 girls who were distributed approximately equally among treatments. The children were randomly assigned to treatments in blocks of four which were based on ranking in terms of chronological age.

Procedure

The pairs consisted of line drawings of familiar objects with no obvious formal and conceptual similarity. The lists involved four pairs in each of List 1 and List 2. The R and W items were randomly determined, with the restrictions applying to identical E and C pairs between sets of lists.

The pairs were presented manually one at a time at an unpaced rate. Children were urged to make a choice if they delayed more than 10 seconds. The pairs were presented to *Ss* who responded by pointing to and verbalizing the R item. The *Ss* assigned to the verbalization treatment pronounced the R and W items on all lists before making a choice. If an incorrect choice was made, the correct item was identified. The serial order of the pairs as well as the left-right position of each item in

a pair was changed in a random fashion on each trial to prevent serial or position learning. The criterion for learning on each list was two successive correct trials.

Prior to List 1, a practice list consisting of two pairs was presented to *Ss* to determine whether the instructions were understood. Following the practice list, *Ss* learned List 1 and List 2, and then relearned List 1. Following practice on List 1 and List 2, the *Ss* were told, "Now I want you to play the same game with some new pictures."

RESULTS

Inspection of the data indicated no obvious differences in difficulty between lists or between pairs assigned to E and C treatments and therefore list differences were omitted for all subsequent analyses.

Acquisition

The mean trials to criterion for the verbalization (V) and non-verbalization (NV) groups were 1.7 and 3.3, respectively, t (18) = 1.68 $p < .20$. The same pronounced trend was evident when the number of errors to criterion was used as the dependent variable. Here the errors on Trial 1 were omitted because performance on this trial was attributable to chance. The mean errors for the V and NV treatments were 1.3 and 5.0, $t = 1.85$, $p < .10$. Although not statistically significant, both dependent measures suggest faster learning when the children were instructed to pronounce the R and W items before making a choice.

Transfer

Learning also occurred rapidly on List 2. Figure 1 provides mean errors for E and C pairs at each of Trials 1–3. As may be seen, positive transfer was obtained; i.e., performance on E pairs exceeded that for C pairs for both V and NV treatments. Two statistical analyses of these data were performed. The first used errors on Trial 1 as the dependent measure while the second involved errors summed over Trials 1–2. The E/C main effect attained statistical significance in each analysis $F(1,18) = 5.05$, $p < .05$ for Trial 1, and $F(1,18) = 7.31$, $p < .025$ for Trials 1–2. Neither the effect of verbalization nor the interaction approached statistical significance in the analyses. It is interesting to note that positive transfer was obtained even though the children were not informed of the relation between List 1 and List 2.

Relearning

The data indicated very rapid relearning with means of 0.1 and 0.7 trials for the V and NV treatments, t (18) = 1.76, $p < .10$. No differ-

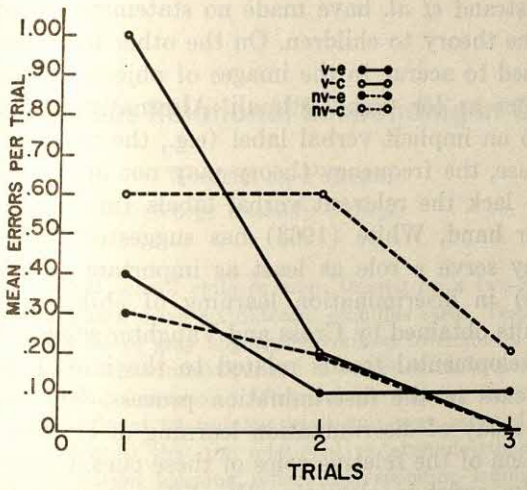


FIG. 1. Mean errors per trial for E and C pairs.

ences were found in relearning between E and C pairs even though the differences in frequency units for R items (accumulated on List 1) and W items (accumulated on List 1 and List 2) could be expected to be less discriminable relative to the point immediately following original learning on List 1.

DISCUSSION

In general, the results do not support the applicability of frequency theory in its present form to discrimination learning in nursery school children. The data indicated a strong trend toward facilitative effects of verbalizing R and W items whereas pronounced inhibitory effects of this variable have been found in VD tasks when college students (Goulet and Hoyer, 1969; Kausler and Sardello, 1967; Sardello and Kausler, 1967) or fifth-grade children (Goulet and Hoyer) have served as *S*s.

In turn, the present transfer data indicated pronounced positive transfer for W_1-R_2 pairs whereas Dean and Kausler, using a VD task and college *S*s, have found inhibitory effects with the same paradigm and under similar instructions to *S*s. Frequency theory does imply that inhibitory effects should be manifest if *S*s responded using Rule 1 during List-2 acquisition. Utilization of Rule 1 is functional on List 1 and would be expected to be transferred to List 2 if *S*s were not informed of the relation between the two lists.

It should be mentioned that the present study was not designed as a direct test of frequency theory. The theory is at present, directly relevant only for VD tasks which involve verbal stimuli (e.g., words).

In addition Ekstrand *et al.* have made no statements regarding the applicability of the theory to children. On the other hand, frequency units could be assumed to accrue to the images of objects (pictures) in much the same manner as for verbal stimuli. Alternatively, frequency units could accrue to an implicit verbal label (e.g., the name of the picture). In the latter case, the frequency theory may not apply because nursery school children lack the relevant verbal labels (mediators).

On the other hand, White (1963) has suggested that negative cues (W items) may serve a role as least as important as that of positive cues (R items) in discrimination learning of children. As mentioned above, the results obtained by Cross and Vaughter suggest the possibility of marked developmental trends related to the importance of positive and negative cues in the discrimination process. The use of transfer designs in the study of discrimination learning of children can assist in the determination of the relative roles of these cues. Comparisons of this nature may also assist in clarifying the different mechanisms involved in discrimination learning and transfer when different stimulus items (objects, pictures, words, etc.) are used. As mentioned above, the transfer data do provide some support for the assumption made by Cross and Vaughter that children approximately 4 years of age respond in discrimination learning on the basis of avoiding the incorrect response.

REFERENCES

Cross, H. A., and Vaughter, R. M. The Moss-Harlow effect in preschool children as a function of age. *Journal of Experimental Child Psychology,* 1966, **4,** 280–284.

Ekstrand, B. R., Wallace, W. P., and Underwood, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review,* 1966, **73,** 566–578.

Goulet, L. R., and Hoyer, W. J. The effects of verbalization on verbal-discrimination learning and associative recall in children and adults. *Journal of Experimental Child Psychology,* 1969, **7,** 434–439.

Kausler, D. H., and Dean, M. G. Direct and mediated transfer for right and wrong responses in verbal discrimination learning. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6,** 672–674.

Kausler, D. H., and Sardello, R. J. Item recall in verbal-discrimination learning as related to proninciation and degree of practice. *Psychonomic Science,* 1967, **7,** 285–286.

Sardello, R. J., and Kausler, D. H. Associative recall in verbal-discrimination learning as related to pronunciation and degree of practice. *Psychonomic Science,* 1967, **8,** 253–254.

Underwood, B. J., Jesse, F., and Ekstrand, B. R. Knowledge of rights and wrongs in verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3,** 183–186.

White, S. H. Learning. *In* H. W. Stevenson (Ed.) *Child Psychology: The Sixty-second Yearbook of the National Society for the Study of Education.* Chicago: University of Chicago Press, 1963. Pp. 196–235.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 485–491 (1969)

Absolute versus Relational Responding in Children

WALTER L. PORTER
George Peabody College[1]

Forty fourth-grade school children were trained on a two-choice brightness discrimination task using six different stimulus cards. Testing was done using the same shades of gray slightly rearranged to enable a response one way for a relational discrimination and a response the other way for an absolute-stimulus discrimination. After initial training and testing the *Ss* received an additional 60 training trials and were retested. The results lend evidence to suggest that the relational vs. absolute-stimulus interpretations of discrimination learning reflect corresponding learning strategies from the *Ss* habit-family hierarchy.

The phenomenon of transposition has been the subject of a number of experimental studies having special interest centered around relational and absolute stimulus theories of discrimination learning (Kimble, 1961). Research has provided evidence that organisms will respond either to the relational or absolute characteristics of the stimuli depending upon the parameters involved. Generally, relational responses will occur when stimuli are near on the stimulus scale (Ehrenfreund, 1952; Hunter, 1953; Stevenson and Bitterman, 1955); when stimuli are presented simultaneously (Baker and Lawrence, 1951; Jackson and Jerome, 1940); and trials are spaced (Rudel, 1955). Moreover, for children there is evidence (Kuenne, 1946; Alberts and Ehrenfreund, 1951) indicating a developmental change in absolute vs. relational responding: the incidence of transposition increases with age.

The tendency to respond to the relational or absolute-stimulus characteristics of a discrimination learning task may represent different strategies within the child's habit-family hierarchy. The present experiment was designed to study the reliability of these strategies—the consistency of absolute vs. relational responding—in a discrimination learning situation.

The discrimination task was modeled after that used by Lawrence and

[1]The author (now at the University of Miami), is indebted to Prof. Norman R. Ellis for his valuable guidance during the study. Gratitude is also expressed to Dr. W. Knox McCharen, Principal of Peabody Demonstration School, for his cooperation in furnishing subjects.

DeRivera (1954). They trained rats on a successive discrimination task using cards with varying shades of gray on the top half and a common intermediate gray on the bottom half. For the three lighter shades of gray a response to the right was correct: for three darker shades of gray a response to the left was correct. Duing testing the seven shades of gray were paired together so that a response to one side was appropriate for a relational discrimination and a response to the other side was appropriate for an absolute discrimination. They found that 65% of the choices were as predicted by relational theory.

METHOD

Materials. The stimulus materials consisted of cards of varying shades of gray adapted from the Lawrence-DeRivera study. Mixtures of flat-black and flat-white paint were used to obtain seven gradations of gray (selected on a purely subjective scale) ranging from white to black (numbered from 1 to 7 respectively). The stimuli used on the learning task consisted of six cards. Each card had one of six shades of gray (1, 2, 3, 5, 6, and 7) on the top half and a common intermediate gray (4) on the bottom half. Thus the cards varied from white over intermediate gray to black over intermediate gray as follows: 1/4, 2/4, 3/4, 5/4, 6/4, 7/4.

During the test trials six additional cards were used with shades of gray as follows: 2/1, 3/1, 3/2, 5/6, 5/7, 6/7. For each trial a card was placed on a small apparatus allowing the *S* to respond to a right or left choice. The apparatus consisted of a vertical plywood face, 12 inches wide by 10 inches high, attached to a plywood base. A window, 3.5 inches wide by 4.75 inches high, was cut out of the center of the vertical facing. Two push-button switches were attached to the base 7 inches apart: one on the right, one on the left. The base extended out 3 inches toward the *S*. Pressing the correct button resulted in the sounding of a bell; which acted as the reinforcement. The apparatus was painted with a clear varnish.

Subjects. The *S*s were drawn from fourth grade classes at the George Peabody Demonstration School. A sample[2] of 40 *S*s, ranging in age from 9 years 4 months to 10 years 3 months, and in IQ from 78 to 150 was used. The IQ information was obtained from their results on the Lorge-Thorndike Intelligence Test which produced a mean IQ of 113 with a SD of 22.

Procedure. Each *S* was invited to sit facing the apparatus at a small table across from *E*; and the following instructions were given:

[2] From an original sample of 44 *S*s, four were unable to reach the training criterion by the end of 25 trial blocks (150 trials).

"You see these buttons. They will ring a bell. Here, try this one (demonstrating *S*'s right button). Okay, now try this one (demonstrating *S*'s left button). Now see this window. This will tell you which button to press. When something is in the window, one of the buttons will ring the bell but the other button won't ring the bell. When it is one way, one of the buttons will ring the bell; when it is the other way, the other button will ring the bell. Remember, what is in the window will tell you which button will ring the bell. Do you want to try to figure it out?"

The six training cards were then presented in random order in the window of the apparatus. For cards 1/4, 2/4, and 3/4 a response to the left button rang the bell. For cards 5/4, 6/4, and 7/4 a response to the right button rang the bell. This allowed the *S* to respond to either the relational properties of the stimuli (top half lighter vs. top half darker) or to their absolute properties (light cards vs. dark cards). In order to eliminate position preference the training cards were presented in blocks of all six cards; randomization being within each block. The training criterion was two consecutive correct blocks (12 trials).

When the training criterion was reached the 12 test cards were presented at random to the *S* who was required to point (no reinforcement was given for any test card) to the button of his choice. The six training cards were presented upside down (4/1, 4/2, 4/3, 4/5, 4/6, 4/7) causing a response based on a relational discrimination to reverse sides while a response based on the absolute discrimination remained the same. The other six testing cards (2/1, 3/1, 3/2, 5/6, 5/7, 6/7) likewise required an opposite response for a relational versus an absolute discrimination.

Thus the procedure allowed the *S* to utilize one of two response strategies. An appropriate response to the left for example, could be elicited whenever one or more of shades 1, 2, or 3 appeared (1/4, 2/4, 3/4 during training; 4/1, 4/2, 4/3, 2/1, 3/1, 3/2 during testing), or whenever the top shade was lighter than the bottom one (1/4, 2/4, 3/4 during training; 4/5, 4/6, 4/7, 5/6, 6/7 during testing): thus representing the absolute and relational strategies, respectively.

After the initial test trials (T-1) the *S* returned to training and continued for ten blocks of trials (60 trials). The *S* was then retested (T-2) using the same procedure.

RESULTS AND DISCUSSION

The results obtained from T-1 and T-2 are illustrated in Figure 1. The figure clearly reveals the bimodal character of the distributions; lending

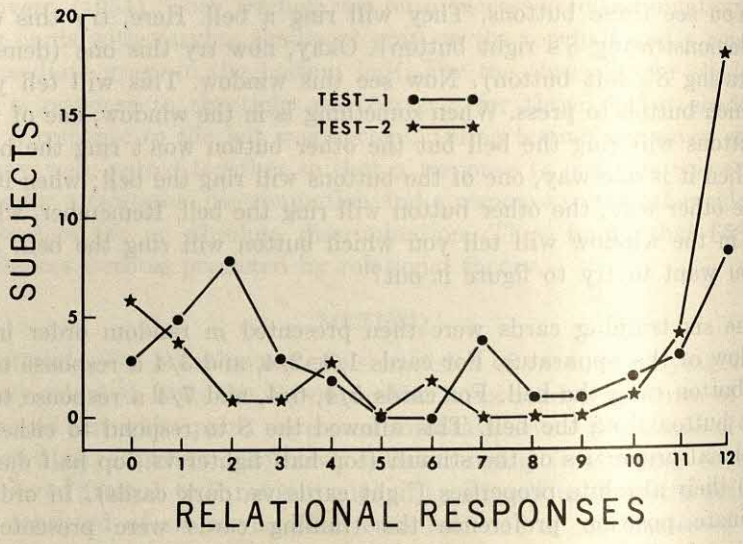


FIG. 1. Number of subjects producing relational responses on the two test series.

support for the contention that the *S*s tended to employ either a relational or an absolute strategy. The middle portion of the distribution (arbitrarily set from 3 to 9 inclusive) represents those *S*s whose test behavior (e.g., position preference, response alternation) was inconsistent with the relational-absolute dichotomy of interest.

Examination of the test series data revealed that the *S* sample could be classified into four response groups (see Table 1) as follows: (1) 13 *S*s (R) who were consistent in responding to the relational qualities of the stimuli (10 to 12 relational responses on each test), (2) nine *S*s (A) who were consistent in responding to the absolute qualities of the stimuli (zero to two relational responses on each test), (3) five *S*s (S) who shifted from a consistent absolute strategy (zero to two relational responses on T-1) to a consistent relational strategy (10 to 12 relational responses on T-2), and (4) 13 *S*s (U) who were unstable or inconsistent on either or both test series (three to nine relational responses). There were no *S*s who shifted from a relational strategy on T-1 to an absolute strategy on T-2. Thus, on T-1 there were 13 *S*s with a relational strategy (group R) and 14 *S*s with an absolute strategy (groups A and S); on T-2 there were 18 *S*s with a relational strategy (groups R and S) and nine *S*s with an absolute strategy (group A). The difference, of course, was a result of the changed strategy of group S.

The initial learning performance of the four response groups is illustrated in Figure 2. Analysis of variance calculations were made comparing

TABLE 1
MEDIANS AND RANGES OF RELATIONAL RESPONSES

Group	N	Test 1	Test 2
Group R	13		
Median		11.69	11.87
Range		10–12	11–12
Group S	5		
Median		1.25	12.00
Range		0–2	12–12
Group A	9		
Median		1.60	0.25
Range		0–2	0–1
Group U	13		
Median		4.25	5.75
Range		0–9	1–12

groups, S, A, and R on both trial-blocks to criterion scores and error scores from the initial training trials data.[3] There were no significant trial-blocks to criterion differences, $F(2/24) = 2.85$. However, there were significant error differences, $F(2/24) = 3.35$, $p < 05$. Further analysis revealed that group A had a significantly higher mean error than group

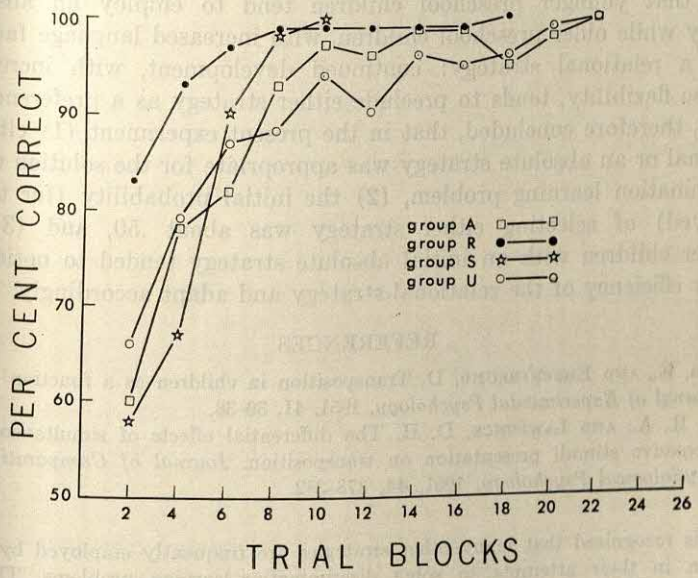


FIG. 2. Initial learning performance for the four response groups.

[3] The excessive variance of the group U data, plus the "residual" nature of the category, led to its exclusion from the analysis of learning behavior.

R. No other differences were significant. Although the trial-blocks to criterion differences were not significant, the significant results of the error analysis together with the appearance of the learning curves suggests that the solution of this particular discrimination task is achieved more efficiently when *S*s use a relational strategy.

The mean IQ scores for the four response groups differed as follows: (1) group R = 117, (2) group A = 112, (3) group S = 133, and (4) group U = 103. An analysis of variance test of these differences produced a significant F ratio, $F(3/36) = 3.09$, $p < .05$. Further analysis revealed that the mean IQ score for group S was significantly greater than both those of groups A and U. No other differences were significant. Thus, considering all of the *S*s with an absolute strategy on T-1, those who changed (group S) to a relational strategy on T-2 had a significantly higher mean IQ than those who did not change (group A). There was no significant IQ difference between the two groups (R and A) whose strategy was consistent from T-1 to T-2.

These results together with those of past research indicate the possibility of developmental changes relative to the use of these strategies (absolute or relational) in the solution of ambiguous discrimination problems such as the one employed in the present experiment.[4] It is suggested that younger preschool children tend to employ an absolute strategy while older preschool children, with increased language facility, prefer a relational strategy: continued development, with increasing response flexibility, tends to preclude either strategy as a preference.

It is, therefore concluded, that in the present experiment (1) either a relational or an absolute strategy was appropriate for the solution of the discrimination learning problem, (2) the initial probability (for the *S*s employed) of selecting either strategy was about .50, and (3) the brighter children with an initial absolute strategy tended to notice the greater efficiency of the relational strategy and adapt accordingly.

REFERENCES

ALBERTS, E., AND EHRENFREUND, D. Transposition in children as a function of age. *Journal of Experimental Psychology,* 1951, **41**, 30–38.

BAKER, R. A., AND LAWRENCE, D. H. The differential effects of simultaneous and successive stimuli presentation on transposition. *Journal of Comparative and Physiological Psychology,* 1951, **44**, 378–382.

[4] It is recognized that many other strategies are frequently employed by young children in their attempts to solve discrimination-learning problems. These of necessity must extinguish, however, if the *S* is to achieve consistent success with the present kind of task involved (those designed to allow success only by means of an absolute or a relational strategy).

EHRENFREUND, D. A. A study of the transposition gradient. *Journal of Experimental Psychology,* 1952, **43,** 81–87.

HUNTER, I. M L. The absolute and relative theories of transposition behavior in rats. *Journal of Comparative and Physiological Psychology,* 1953, **46,** 493–497.

JACKSON, T. A., AND JEROME, E. Studies in the transposition of learning by children: IV, A preliminary study of patternedness in discrimination learning. *Journal of Experimental Psychology,* 1940, **26,** 432–439.

KIMBLE, G. A. *Hilgard and Marquis' Conditioning and Learning.* Appleton-Century-Crofts: New York, 1961.

KUENNE, M. R. Experimental investigation of the relation of language to transposition behavior in young children. *Journal of Experimental Psychology,* 1946, **36,** 471–490.

LAWRENCE, D. H., AND DERIVERA, J. Evidence for relational transposition. *Journal of Comparative and Physiological Psychology,* 1954, **47,** 465–471.

RUDEL, R. G. Transposition of response by children trained in intermediate-size problems. *Journal of Comparative and Physiological Psychology,* 1957, **50,** 292–295.

STEVENSON, H. W., AND BITTERMAN, M. E. The distance effect in the transposition of intermediate size by children. *American Journal of Psychology,* 1955, **68,** 274–279.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 492–513 (1969)

Facilitation of Repeated Free Operant Discrimination Reversals[1]

RICHARD P. TOISTER
University of Miami

AND

J. S. BIRNBRAUER
University of North Carolina at Chapel Hill

Two experiments are reported which studied the effects of several parameters on responses after discrimination reversals, "errors," by retarded subjects. In Exp. 1 fixed-ratio and variable-ratio schedules of reinforcement and access to a switching key were investigated. In Exp. 2 the effects of a schedule change (FR 25 to FR 1) prior to reversal and a "count" of four reinforcements to reversal were studied with a design which permitted determining the effects of both variables simultaneously. Both experiments showed the diminution of errors after repeated reversals that has been reported in other studies of discrimination reversals and "learning set." Access to a switching key resulted in reliably fewer errors in three of four subjects but many errors in excess of the ratios nevertheless were emitted. The number of responses after reversals may exceed FR schedules in direct relationship to the value of the FR—the higher the FR, the greater the number of responses in excess of the ratio. Schedule change prior to reversals and regularity in programming reversals resulted in errorless switching.

If in a simple discrimination task the S^D and S^Δ are repeatedly reversed, a common finding is a decrease in "errors" (responses to the just previous S^D) to some asymptotic value. This finding has been reported in both discrete trial (Longnecker, 1964; Plenderleith, 1956; Harlow,

[1]Experiment 1 was supported in part by NIMH Research Grant M2232 (Dr. S. W. Bijou, Principal Investigator) and was conducted at Rainier School, Buckley, Washington, while the second author was on the staff of the University of Washington. Thanks are due S. W. Bijou, Robert Orlando, Russell M. Tyler, and several faithful research assistants who don't wish to be reminded of the many hours spent. Experiment 2 was submitted in partial fulfillment of the senior author's Ph.D. requirements at the University of Miami and was supported in part by NIMH project grant MH 14914 and conducted at Murdoch Center, Butner, North Carolina. The contributions of Drs. Marshall R. Jones, Robert Allen, and Daniel Cruse are gratefully acknowledged. The authors would also like to thank Jack Sandler and Edward Malagodi for their helpful suggestions.

1959) and free operant studies (Deterline, 1960; Long, 1963). One variable affecting the level of the asymptotic value of errors is, of course, the schedule of reinforcement. For example, an FR 10 schedule requires ten responses after a reversal before the new contingencies are contacted. If *S* switches and responds to the previous S^{Δ} after only ten responses, an errorless reversal is described. (While the terms "reversal" and "switch" are often used interchangeably, the authors will use "reversal" to describe a change in contingencies and "switch" to describe *Ss*' behavior.)

The purposes of the present studies were to analyze some variables influencing the asymptotic level of errors after experiencing many reversals and to effect errorless discrimination reversals in a free operant situation. While there have been numerous studies using a discrete trial design (Ellis, 1958; Girardeau, 1959; House and Zeaman, 1959; Mackintosh, 1965; Shepard, 1957; Sperling, 1965) few systematic free operant investigations have been reported.

EXPERIMENT 1

One objective of this study was to determine the effects of giving *S* the opportunity to choose discriminative stimuli by pressing a "choice key." When the "choice key" was available, *S* could switch from S^{Δ} to S^{D} or vice versa at any time. When it was not accessible, *S* had to "wait out" S^{Δ} periods. The second objective was to determine the relationship between the size of the fixed-ratio schedule of reinforcement (FR 15, 38, and 100) and the number of responses emitted before switching after reversal. The third objective was to compare responses before switching when the schedule of reinforcement was a variable ratio (VR) with that when FR was in effect.

Method

Subjects

The four subjects described in Table 1 were selected from the trainable and educable groups at Rainier School, Buckley, Washington. The

TABLE 1
CHARACTERISTICS OF *Ss*

S	Sex	CA	IQ	Test
GM	M	18	60	WAIS
MT	F	25	55	PPVT
JS	F	23	35	PPVT
BS	M	17	61	PPVT

Ss manifested no visual or motoric handicaps that would interfere with discriminating the red and green stimuli or operating the response lever or key, and were, by reputation, reliable and cooperative.

Apparatus

The study was conducted in a 7×7 foot sound-controlled room. The experimental equipment consisted of: (1) an easily operated chrome lever made from a squeeze mop handle; (2) a reinforcer chute and catcher to the right of the lever; (3) a white jewel light mounted on the reinforcer chute; and (4) a telegraph key to the left of the lever. A one-way glass and a 7×13 inch milkglass panel were mounted on the wall directly above the level. Relay circuitry and response recording equipment were located in an adjacent room. [The laboratory is described in more detail in Orlando and Bijou (1960); Orlando, Bijou, Tyler, and Marshall (1960).]

The jewel light was illuminated at the instant of reinforcement and stayed on for the 1.5 seconds that reinforcer delivery required. The reinforcers at first were pennies, assorted candies (M & M's, candy corn, jelly beans) or a mixture. Later, beads were delivered during the session and these were exchanged at the end of each session for pennies, candy, trinkets, or any combination *S* selected.

Red and green lights were mounted behind the milkglass panel. These were correlated with reinforcement (S^D) and extinction (S^Δ). The panel appeared gray when neither light was on.

Operation of the telegraph key (referred to as the choice key) moved the stepping relay which controlled the color of the panel one position. From one to three presses of the choice key were required to change from green to red or vice versa. During "no choice" phases a gray box was screwed in place over the choice key to make it inaccessible.

Procedure

At the beginning of the first session the choice key was not accessible and the panel light was the color of S^D for that *S*. The *E* seated *S* before the equipment and gave instructions and a demonstration that pressing the lever would produce candy (pennies). The *E* pressed the lever five times at a rate of approximately two responses per second, took the reinforcer and said, "Now you try it." After the subject had received one reinforcement on FR 5, *E* left the room with the comment, "Get lots of candy (pennies) and I'll be back when it's time to go." On subsequent sessions *E* repeated the final comment. Nothing was said about reversals or schedule changes.

When the choice key was first uncovered, *S* was shown that it con-

trolled the lights and was told that he could use it "if you don't like the color of the lights." Nothing was said about its inability to influence the gray condition and the instructions were not intended to either encourage or discourage choice responses. The no-choice condition was reinstated without comment by replacing the box covering the key.

The first step in the procedure was to teach *S* not to respond during S^{Δ}. This was done after the "rapid acquisition" procedures described by Bijou and Orlando (1961). The procedure consisted of reinforcing the subject three to five times in the presence of S^{D} and then changing to S^{Δ}. S^{Δ} remained in effect until *S* stopped responding for a required length of time (e.g., 5 seconds). Then, S^{D} was presented. The pause in responding required in S^{Δ} was gradually extended and the cycle continued until *S* stopped responding with the onset of S^{Δ} but was responding steadily during S^{D}. Meanwhile, the schedule of reinforcement was increased gradually to the initial value selected for study. Finally, the time controlled S^{D}–S^{Δ} sequence began. (One *S*, JS did not respond to the rapid aquisition procedure. During her first 25 sessions she sporadically emitted responses to S^{Δ}. Consequently, the choice key was made accessible as a last effort to improve her accuracy.)

The sequence of conditions was 90 seconds of S^{D} or S^{Δ} alternating with 15 seconds of gray light. In the terminology of Ferster and Skinner (1957) the baseline was a time-controlled Mult FR or VR (S^{D}) TO Ext (S^{Δ}). The order of S^{D} and S^{Δ} was random except that S^{D} was always in effect when *S* entered the laboratory and neither S^{D} nor S^{Δ} could occur more than three times in succession. The session duration varied from 20 to 40 minutes depending upon *S*'s performance. Similarly, the type of reinforcement was changed when this step appeared necessary to maintain steady responding.

After *S* acquired the initial discrimination, reinforcement was discontinued in S^{D} and programmed during S^{Δ} without warning or disruption of the light sequence. By flipping a switch the stimulus panel light that signified S^{D} was correlated with extinction and the light that signified extinction became correlated with reinforcement. Thereafter, reversals were presented repeatedly. The number of response-free S^{Δ} periods required between reversals was varied irregularly from two to six with occasional longer periods. The minimum criteria for reversal were: (1) responding steadily after the onset of S^{D} and (2) no lever responses in S^{Δ} for two consecutive S^{Δ} periods. Reversals were programmed so as to prevent *S*s from coming under the control of the schedule of reversals and to obtain a situation in which nonreinforcement was the only cue that a contingency change had taken place.

Responses emitted during the first 90-second S^{Δ} period after reversal

TABLE 2
SEQUENCES OF CONDITIONS AND RESULTS

Subject schedule	Sessions-reversals	Responses to switch: Last reversals	Responses to switch: Median	Responses to switch: Range
GM				
FR 38 NC	30–25	7	88	72–108
VR 38 NC	23–27	7	196	88–328
FR 38 NC	10–10	7	124	92–276
FR 38	16–45	7	88	56–108
FR 38 NC	10–07	7	132	84–192
FR 38	04–07	7	84	60–108
FR 38 NC	13–12	7	120	84–144
FR 38	02–06	6	80	52–108
FR 100	11–17	7	199	100–324
Conditions omitted	—	—	—	—
FR 15	05–21	7	63	45–132
MT				
FR 38 NC	08–13	7	136	108–252
FR 38	15–24	7	96	56–152
FR 38 NC	03–06	6	212	120–300
FR 38	04–06	6	54	36–068
FR 38 NC	08–12	7	116	84–196
FR 38	04–07	7	68	42–164
Conditions omitted	—	—	—	—
FR 100	06–11	7	192	130–293
JS				
FR 100	35–41	7	352	328–364
FR 38	19–49	7	112	80–160
FR 38 NC[a]	10–05	5	212	188–240
FR 38	03–10	7	116	96–128
FR 38 NC	08–04	4	154	136–232
FR 38	03–07	7	92	68–104
VR 38	06–13	7	120	76–264
FR 38	06–13	7	112	76–140
VR 38	07–14	7	136	112–208
FR 38	11–28	7	104	92–155
FR 100	10–19	7	279	172–384
FR 38	08–15	7	126	96–152
VR 38	06–11	7	162	104–220
BS				
VR 38 NC	15–18	7	192	104–264
FR 38 NC	11–25	7	108	92–200
VR 38 NC	04–06	6	122	100–140
FR 100 NC	08–13	7	260	240–280
FR 38 NC	07–12	7	116	84–156
FR 38	07–11	7	64	48–232

TABLE 2 (*Continued*)

Subject schedule	Sessions-reversals	Responses to switch		
		Last reversals	Median	Range
FR 38 NC	02–04	4	64	60–072
FR 38	06–11	7	67	52–144
VR 38	12–23	7	84	58–132
FR 38	08–11	7	65	58–100
Conditions omitted	—	—	—	—
FR 100 NC	07–12	7	211	151–309
FR 100	06–20	7	196	154–217
FR 100 NC	03–11	7	163	126–296

[a] Choice key not accessible.

were the data of focus and are referred to as "responses to switch." (Several examples appear in the cumulative records in Figures 1, 2, and 3. Responses to switch are those emitted in the interval below the reversal number.) Responses in the first interval seemed to provide the least confounded measure of sensitivity to the schedule of reinforcement.

When the number of responses to switch stabilized; i.e., ceased to show an increasing or decreasing trend, the experimental conditions were changed. These remained in effect until S's responses to switch leveled off; then other conditions were instated. Each *S* was his own control and within-*S* replication of effects was the aim (Sidman, 1960).

The schedules tested which yielded consistent and reliable data and therefore will be reported were: FR 15, FR 38, VR 38 (range 4–72), and FR 100. Table 2 shows the sequence of conditions for each *S*. Since the choice key was present most of the time, schedules with the choice key absent are designated NC ("no choice"). The order of schedules differed among *S*s, not all *S*s contacted all conditions and there were wide variations in the number of reversals and sessions per condition. These variations reflect the fact that changes in conditions were contingent upon *S*'s performance. In general, high numbers of reversals reflect instability, a persisting trend or an effort to determine if prolonged contact with a schedule would have an effect. Low numbers mean that a change in conditions had no effect or a change reinstated performance similar to that obtained earlier under the same schedule.

RESULTS

Table 2 summarizes the findings. The results of the main comparisons will be presented after some remarks about the performances the various

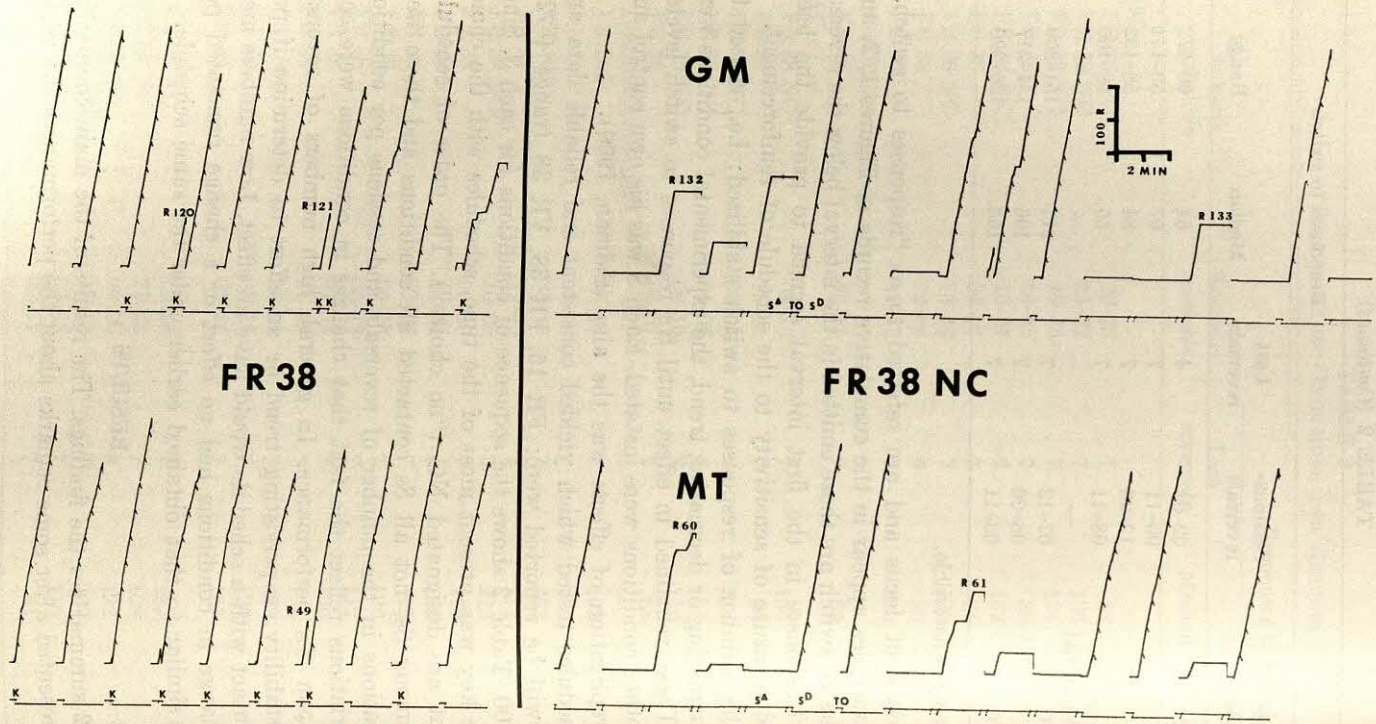


FIG. 1. Stable performances on multiple FR 38 TO EXT with and without access to the choice key. NC = choice key inaccessible; R = reversal; K = choice key operated; base pen in down position = S^D; base pen in up position = S^A or TO. The cumulative response pen resets with change of components.

schedules generated and the effects of changes in conditions. This should help clarify the design and procedures.

The cumulative records in Figures 1, 2, and 3 exemplify general characteristics of the *Ss'* multiple schedule performance after experiencing several reversals. Response *rates* in S^D were generally high and steady and were not affected by reversals (Figures 1, 2, and 3, R), schedule changes (Figure 2, record 112, and Figure 3, records 35 and 56) and the removal of the choice key (Figure 2, record 54). TO control also was not affected by changes or reversals and responding during the 15-second time-out periods was very unusual. (Figure 1, GM's right record, includes one of the exceptions.)

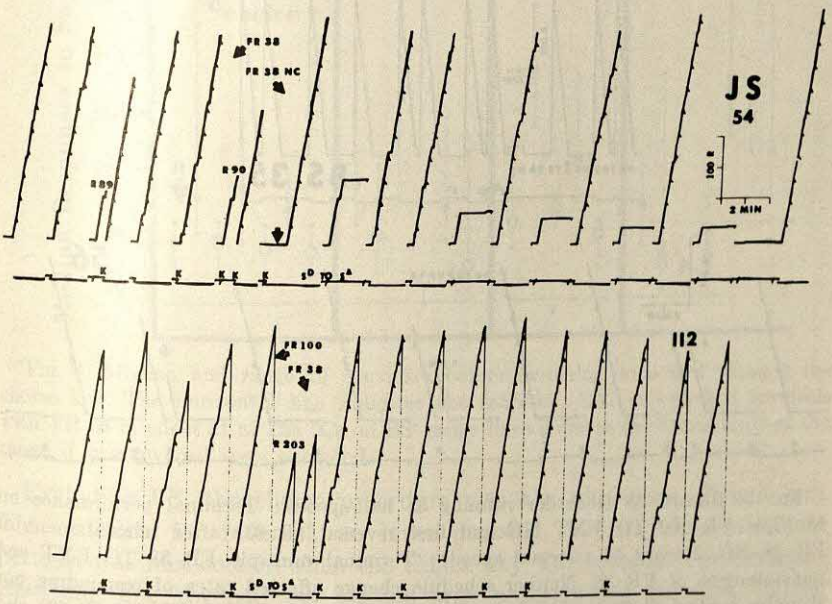


FIG. 2. Record 54: The short-term effect of change from multiple FR 38 TO EXT to FR 38 TO EXT NC. JS was removed from the laboratory while the choice key was covered at the arrow. Record 112: Perfect avoidance of S^Δ and steady S^D responding unaffected by change from FR 100 to FR 38 at R 203. This record characterizes the performance of JS, MT, and GM with the choice key.

Figure 1 shows final performances of GM and MT on FR 38 with the choice key present and with it absent (FR 38 NC). These *Ss* and JS (Figure 2) consistently emitted choice responses very quickly after S^Δ appeared, thereby "keeping themselves in" S^D and made "errors" only when the contingencies were reversed. BS, in contrast (Figure 3, record 56) maintained alternation of S^D and S^Δ by emitting choice responses. He, too, made "errors" infrequently.

Figure 2, record 54, FR 38 NC, shows the high incidence of S^{Δ} responding after reversal found in the records of all *Ss* at first. JS was unusual in that S^{Δ} responding persisted during NC. The record also shows that S^{Δ} rates like S^{D} rates were high, steady, and affected little by changes. Response "runs" with abrupt cessation are typical of extinction after FR reinforcement; the failure of variables to influence rates maintained by ratio schedules is also not uncommon (Ferster and Skinner, 1957).

The effects that emerged were seen in the total number of lever responses emitted in S^{Δ} periods. Hence, the results are in terms of responses to switch.

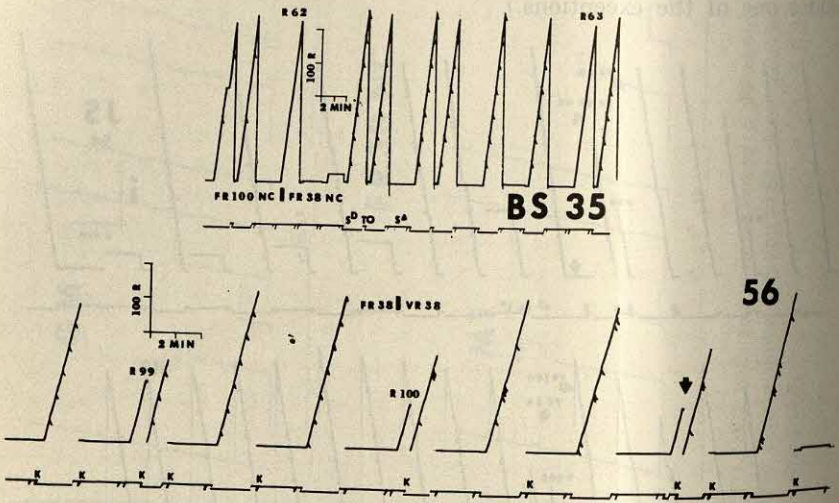


FIG. 3. Record 35 (recorder running at half-speed): Terminal performance on Multiple FR 100 TO EXT NC and first reversal (R 63) after reinstatement of FR 38 NC. Record 56 (normal speed): Terminal multiple FR 38 TO EXT and reinstatement of VR 38. Neither schedule change affected rates of responding nor stimulus control. R 63 produced responses to switch identical to FR 100 (R 62) and two to three times higher than the stable FR 38 level exemplified by R 99 and R 100. Record 56 also shows BS′ emission of choice responses (K) to maintain alternation of S^{D} and S^{Δ}. Note the similarity between the responses to switch before BS "corrected himself" at the arrow and those at R 99 and R 100.

Influence of Choice Key

Figure 4 summarizes the performance of each *S* when the choice key was accessible and when it was not (NC). Three of the *Ss* consistently emitted more responses after reversals and tended to be more variable in the NC condition. Although the opportunity to escape S^{Δ} and restore S^{D} improved performance, the *Ss* generally emitted considerably more responses to switch than the 38 required to contact nonreinforcement.

Note that after contact with VR 38 (Figure 4a) GM's FR 38 NC responses to switch increased and the choice procedure restored performance to its original level.

The fourth *S*, BS, with FR 38 in effect, showed greater variability when the choice key was present; essentially the same results were obtained later with FR 100. BS acquired a chain of switching from S^D after an S^D period and switching from S^Δ after an S^Δ period as seen in Figure 3, record 56.

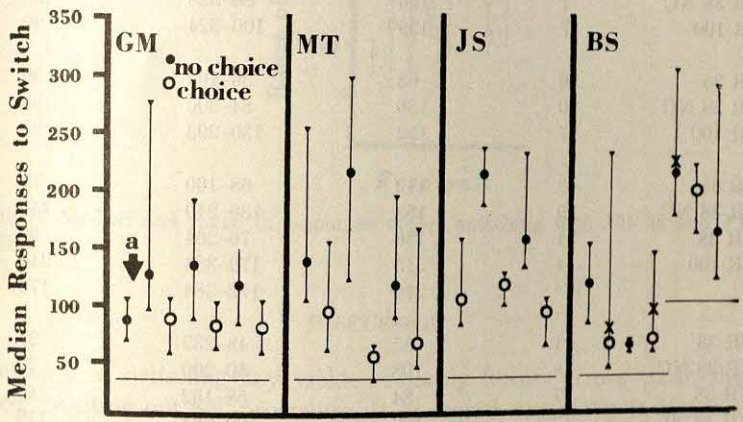


FIG. 4. Median and range of responses before switching with and without the choice key. The horizontal line indicates the schedule. GM experienced reversals with VR 38 in effect at a. The X's on BS' range lines indicate the upper limit of the range if one reversal were excluded.

Excluding BS, the choice procedure was also the more efficient one. The criterion for reversal, no lever pressing for two consecutive S^Δ periods, was met sooner when the choice key was present. For example, less than one reversal per session was programmed for GM during FR 38 NC; from two to three reversals were programmed when the choice key was present (Table 2, GM). JS, during FR 38 NC, met the criterion so infrequently that these conditions were discontinued with only a few reversals each time the condition was presented. With the choice key present, stimulus control of JS' behavior was perfect (Figure 2). The emission of a choice response in the presence of S^Δ is, of course, as good an indicator of a discrimination as "waiting out" S^Δ.

Influence of Schedule Value

The differences between the overall medians on each schedule and the value of the fixed ratio or the mean value of the variable ratio appear in Table 3. Without exception the number of responses to switch in

TABLE 3
DIFFERENCES BETWEEN RATIOS AND OVERALL MEDIAN RESPONSES TO SWITCH

Subject schedule	No. of reversals	Median	Range	Median-ratio difference
GM				
FR 15	7	63	45–132	48
FR 38	20	84	52–108	46
FR 38 NC	28	122	72–276	84
VR 38 NC	7	196	88–328	158
FR 100	7	199	100–324	99
MT				
FR 38	20	68	36–164	30
FR 38 NC	20	136	84–300	98
FR 100	7	192	130–293	92
JS				
FR 38	42	112	68–160	74
FR 38 NC	9	183	136–240	145
VR 38	21	136	76–264	98
FR 100	14	315	172–384	215
	(7[a]	279	172–384	179)
BS				
FR 38	21	65	48–232	27
FR 38 NC	18	108	60–200	70
VR 38	7	84	58–132	46
VR 38 NC	13	157	100–264	119
FR 100	7	196	154–217	96
FR 100 NC	21	211	126–309	111
	(14[a]	187	126–309	87)

[a] First contact with FR 100 excluded.

excess of the ratio was higher for FR 100 than FR 38. GM, however, did not more closely approximate FR 15 than he did FR 38.

Without exception the median number of responses to switch on VR 38 was greater than on FR 38 with JS yielding the most reliable differences. Figure 5 shows that she responded after reversals more often and more unpredictably with VR 38 in effect. Confidence in the reliability of the effect is enhanced by the fact that the last two measurements were taken after extended contact with FR 100. While the other *S*s showed the same relationship between VR and FR, the differences were not as marked. There was considerable overlap in the ranges of responses to switch. Note also in Table 2 that BS' final responses to switch on VR 38 were quite similar to his FR 38 performance.

Table 3 reiterates that NC produced more responses to switch than the corresponding choice condition. In fact, fixed-ratios with no choice generated more responses to switch than the corresponding variable ratios when the choice key was present.

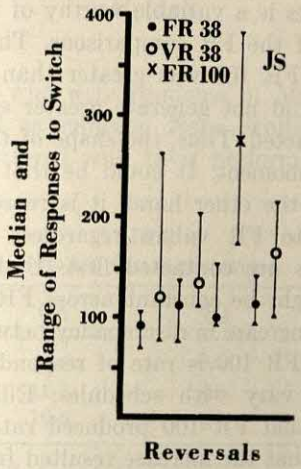


FIG. 5. Median and range of responses before switching with FR 38 and VR 38 for JS.

DISCUSSION

Access to a switching key yielded reliably fewer "errors" than the no choice, "waiting out S^{Δ}" procedure. It appeared, especially in the case of JS and clearly not in BS' case, that teaching *S* to do something in extinction enhanced *S*'s "ability to discriminate" visual stimuli. Switching out of S^{Δ} into S^{D} and not vice versa, by definition, indicates that with each reversal the stimulus discriminative for extinction became a negative reinforcer and that the stimulus discriminative for reinforcement, a positive reinforcer. Access to the choice key, however, did not result in similarly perfect discrimination of the number of responses to reinforcement.

Although variable ratio schedules of reinforcement generally resulted in higher and more variable responses to switch, the differences between VR and FR were often not large and diminished with repeated exposures. It also seemed that schedule effects endured for some time after the schedule had been replaced by another. Thus, a definitive answer will require even longer exposures to a given schedule than were provided in this study and/or returning *S*s to a constant level of performance between exposures to the schedules of primary interest. For example, one might separate schedules with another which is quite different, such as FR 100, between FR 25 and VR 25. The suggestion being made is that VR 25 after FR 25 may generate a different steady state than VR 25 not immediately preceded by FR 25. The range of the VR, of course, must be taken into consideration.

The order of schedules is a variable worthy of manipulation also with regard to the results of the FR comparisons. The number of responses to switch in excess of FR 100 was greater than the number exceeding FR 38 but the latter did not generate greater excess than FR 15, the last schedule GM contacted. Thus, the shape of the function is a matter for conjecture at the moment. It could be that the function is direct above some point. On the other hand, it is reasonable to suppose that closer approximation to FR values regardless of magnitude will be obtained if lower FR's are contacted first. Under these circumstances the excess actually might be constant across FR's. Another factor that might account for the increase in discrepancy between responses to switch with FR 38 and with FR 100 is rate of responding. The rates of these Ss, however, did not vary with schedules. Ellis, Barnett, and Pryer (1960) also reported that FR 100 produced rates similar to FR 25. A second explanation is that the increase resulted from fewer contacts with reinforcement on FR 100 than FR 38. That is, the difference might be eliminated by equating the number of reinforcements on each schedule. This possibility cannot be ruled out since reinforcements were not equal, but the likelihood is diminished by the fact that schedule changes were based on performance rather than time-on-schedule.

A variable which should have considerable bearing on the generality of these results to other humans is the *S*'s experience with counting. Although the present *S*s were sensitive to the relationship between the number of lever presses and the occurrence of reinforcement, if they were counting responses, they were doing so quite inaccurately. JS, whose errors exceeded the ratios most, could not count at all.

EXPERIMENT 2

In view of the greater efficiency of the choice key condition in Exp. 1 and other consistent results (Oatley, Bryant, and Tinson, 1965) a simultaneous or concurrent operant procedure was followed in Exp. 2. Also, one experimental manipulation was applied to reversals from one color to the other (i.e., green-to-red reversals) while another experimental manipulation was applied to reversals from the second color to the first (i.e., from red to green). This design change permitted the development of two baselines (red-to-green reversals and green-to-red reversals).

One objective was to determine the effects of a fixed number of reinforcements prior to a reversal upon switching ("count"). The second was to see if a schedule change, specifically from FR 25 to FR 1 just prior to reversal, could function as a discriminative stimulus for switching. The design permitted both of these questions to be answered at the same time.

Method

Subjects

Three retarded boys who were residents of Murdoch Center, Butner, North Carolina, served as subjects. None exhibited sensory difficulties which would have interfered with their performance. Table 4 presents *S* characteristics.

TABLE 4
SUBJECT CHARACTERISTICS

Subject	CA	IQ	Test
M-1	12-2	58	PPVT
M-2	11-2	65	PPVT
M-3	10-6	81	Stanford-Binet

Experimental Setting and Apparatus

The experimental setting consisted of a subject room and adjacent control room similar to that described in Exp. 1. The subject's room housed two panels separated by a reinforcement delivery tray. Each panel consisted of a translucent glass window above a Lindsley Manipulandum. Each window covered a red and green 25-watt bulb which served the same function as in Exp. 1 and, depending upon the contingencies, the panel was either green or red. Reinforcers were small brass coins which were traded for M & M candies at the end of each session on approximately a 3-coin to 1-M & M ratio.

Procedure

Both red and green lights were present, one color on each panel, at all times except after reinforcement. Responses to the bar under the S^{Δ} light were not reinforced, while bar pulls under the S^{D} delivered a coin on FR 25. Five seconds after reiiforcement, the panel lights came on again with the same color on the same panel or on the alternate panel. The position of red and green lights randomly alternated from the left to the right panel.

The experimental phases were as follows:

Phase I. In the first session, *S* was told to pull the bar under the red light and then the bar under the green light. Responses to the green light were reinforced with a coin on FR 1. After delivery of the first coin, *S* was told that this was like "money" and he could buy M & M candies at the end of the session.

Phase II. After *S* was pulling only the bar under the green light, the reinforcement schedule was gradually increased from FR 1 to FR 25.

Phase III. Reversal baseline. After eight reinforcements on FR 25 to the green light with no S^{Δ} responses, the stimuli were reversed so that red was correlated with FR 25 reinforcement. After the first reversal, the stimuli were repeatedly reversed after four, five, six, seven, or eight reinforcements randomly programmed. If *S* responded more than ten times to the S^{Δ} during a sequence of, for example, six reinforcements, the criterion number of reinforcements to reversal was reset to six.

Phase IV. The FR 25 schedule on red S^{D} became a mixed FR 25 FR 1 with the FR 1 programmed for three reinforcements prior to reversal. When S^{D} was red, the subject received three coins in succession (FR 1) after criteria for reversal were met on FR 25. The five-second TO period followed each coin delivery as usual. Then the stimuli were reversed.

Phase V. After a new baseline appeared, reversals from green to red (G–R) were programmed to occur after every fourth reinforcement on FR 25. Thus, the discriminative stimulus for red-to-green (R–G) reversals could be the interpolated FR 1 for three reinforcements (Phase IV) and the discriminative stimulus on G–R reversals could be the "count" (Phase V) of four coins.

Phase VI. The FR 1 and "count" were removed. (Same as Phase III.)

Phase VII. The FR 1 on R–G reversals and the "count" on G–R were reintroduced. (Same as Phase V.)

Following Sidman's suggestion (1960), all experimental changes were initiated within an experimental session to control for any intersession variability and/or initial "warm-up" effects.

RESULTS

Development of Reversal Baseline

One *S* reached the criterion of eight consecutive reinforcements on FR 25 with no responses to S^{Δ} in the first session. The other two *S*s (M2 and M3) alternated between the S^{D} lever and S^{Δ} lever before responding only to the S^{D} lever. M3 ceased alternating during the fourth session without any intervention; M2 ceased this behavior after *E* entered the room and said, "What color light is that?", each time M2 pulled the bar.

All *S*s made a large number of errors after the first reversal, but achieved a low error rate rather quickly. Records A, D, E, and G of Figure 6 show responding after the first stimulus reversal. (In the figures all responses moved the cumulative response pen; S^{Δ} responses are recorded on the base or event line.) M1 and M3 (Records A, G switched and began responding to the new S^{D} with very little disruption in rate, whereas, M2 (D and E) switched after long periods of no responding.

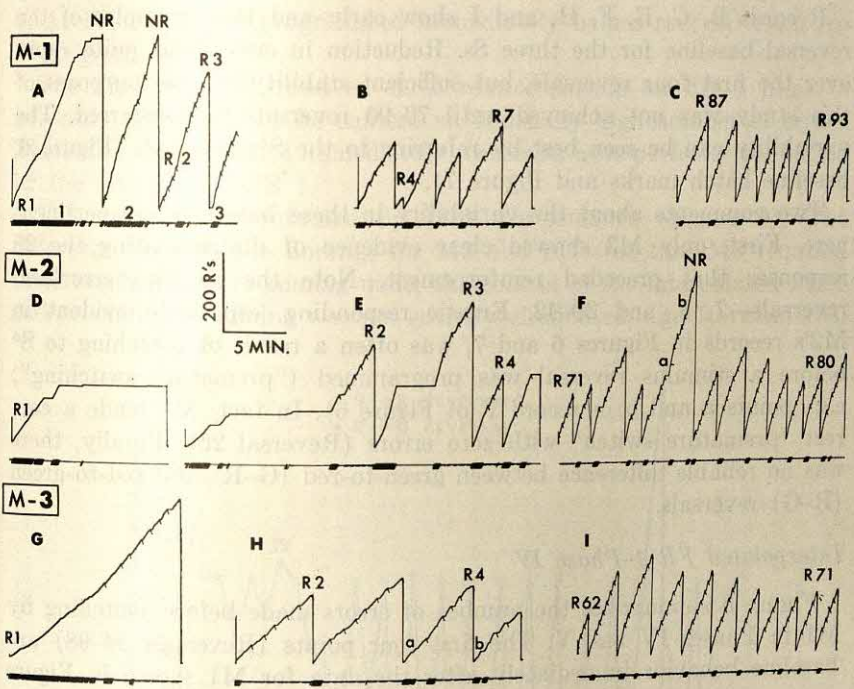


FIG. 6. Representative cumulative recordings for early reversals and the reversal baseline. All responses moved the cumulative recording pen. Event pen marks show S^{Δ} responses. Each pen reset, unless marked NR, indicates a reversal (as do reversal numbers R1, R2, also). Records A, B, D, E, G, and H show early reversal behavior; records C, F, and I show later and final reversal baseline performance for the three subjects.

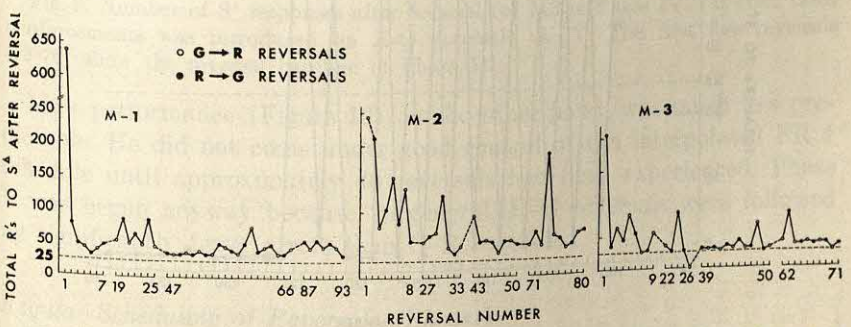


FIG. 7. Total number of S^{Δ} responses after reversals for M1, M2, and M3 in Phase III. The dotted line at 25 indicates the number of responses necessary to contact reversal on the FR 25 schedule.

Records B, C, E, F, H, and I show early and late examples of the reversal baseline for the three *S*s. Reduction in errors was quite rapid over the first four reversals, but sufficient stability for the purposes of this study was not achieved until 70–90 reversals had occurred. The variability can be seen best by referring to the S^{Δ} responses (Figure 6, baseline hatch marks and Figure 7).

Two comments about the variability in these baselines are pertinent here. First, only M3 showed clear evidence of discriminating the 25 responses that preceded reinforcement. Note the nearly "errorless" reversals—7, 8, and 39–42. Erratic responding, especially evident in M2's records in Figures 6 and 7, was often a result of switching to S^{Δ} before a stimulus reversal was programmed ("premature switching"; e.g., points a and b in record F of Figure 6). In fact, M3 made a correct "premature switch" with zero errors (Reversal 26). Finally, there was no reliable difference between green-to-red (G–R) and red-to-green (R–G) reversals.

Interpolated FR 1–Phase IV

Figure 8 summarizes the number of errors made before switching by M1 in Phases IV and V. The first four points (Reversals 94–98) are baseline behavior immediately after the data for M1 shown in Figure

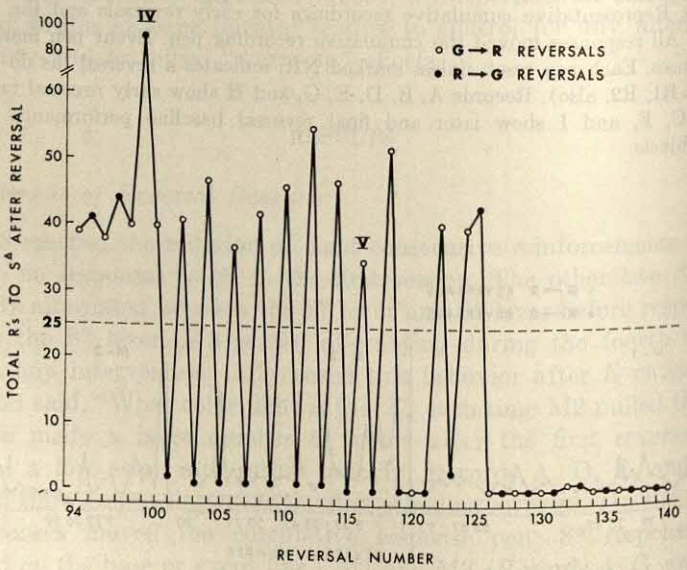


FIG. 8. Number of S^{Δ} responses after reversal for M1 in Phases IV and V. At IV FR 1 for three reinforcements was introduced prior to R-G reversals and at V reversing after every fourth reinforcement was introduced on G-R reversals. The first five reversals (94–98) show the Phase III reversal baseline.

7. At IV, FR 1 was programmed immediately before reversal from red S^D to green S^D.

The first time M1 received three reinforcements on FR 1 prior to reversal (Figure 8, IV) he emitted an unusually high number of errors. Thereafter, M1 switched immediately; reversals were perfectly controlled by the interpolated FR 1.

This result was confirmed by M3's performance (Figure 9). The difference between the findings for M3 and M1 was that M3 required three reversals before coming under the control of the interpolated FR 1 and premature switching was not entirely eliminated, e.g., Reversal 103.

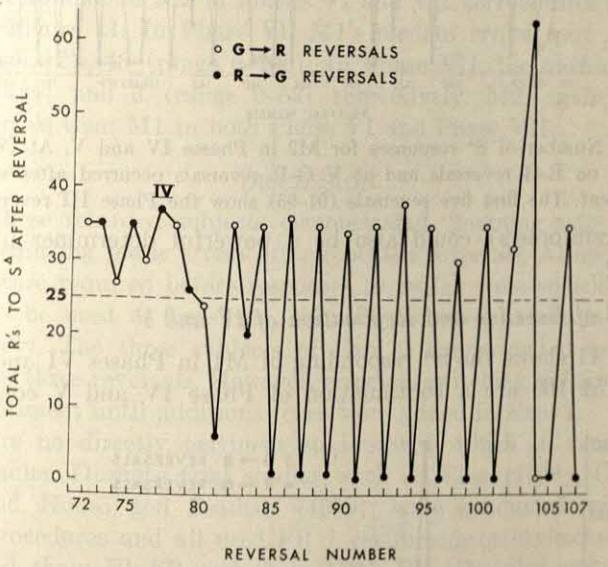


FIG. 9. Number of S^A responses after reversal for M3 in Phase IV. FR 1 for three reinforcements was introduced on R-G reversals at IV. The first five reversals (72–76) show the reversal baseline of Phase III.

M2's performance (Figure 10) on the other hand, was much less predictable. He did not come under good control of the interpolated FR 1 schedule until approximately 40 reversals had been experienced. Phase V was begun anyway because, in general, R–G reversals were followed by significantly fewer errors than G–R reversals.

Regular Scheduling of Reversals—Phase V

This change was made at V in Figures 8 and 10.[2] Both *Ss* demonstrated by their performance that regularity of scheduling a reversal (a "count"

[2] M3 left the institution prior to Phase V.

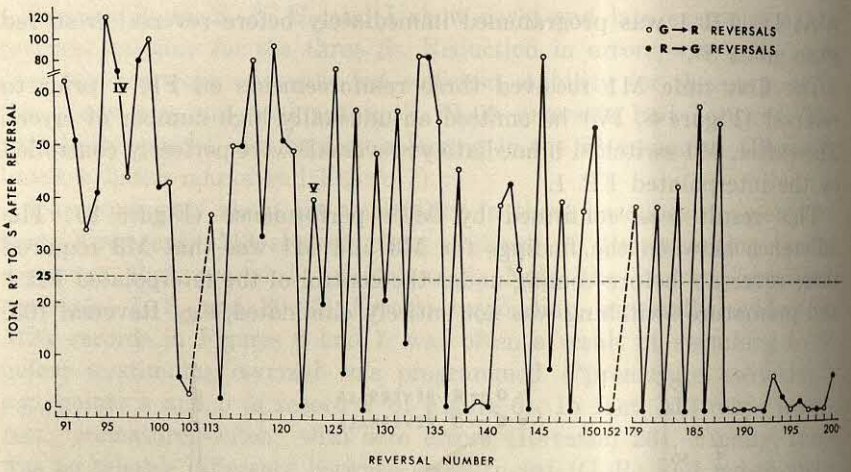


FIG. 10. Number of S^{Δ} responses for M2 in Phases IV and V. At IV FR 1 was introduced on R-G reversals and at V G-R reversals occurred after every fourth reinforcement. The first five reversals (91–95) show the Phase III reversal baseline.

of four reinforcers) could also be a powerful determiner of errorless switching.

Retrieval of Baseline and Replication of IV and V

Figure 11 shows the S^{Δ} responding of M1 in Phases VI and VII. Reversals 154–160 are a continuation of Phase IV and V contingencies.

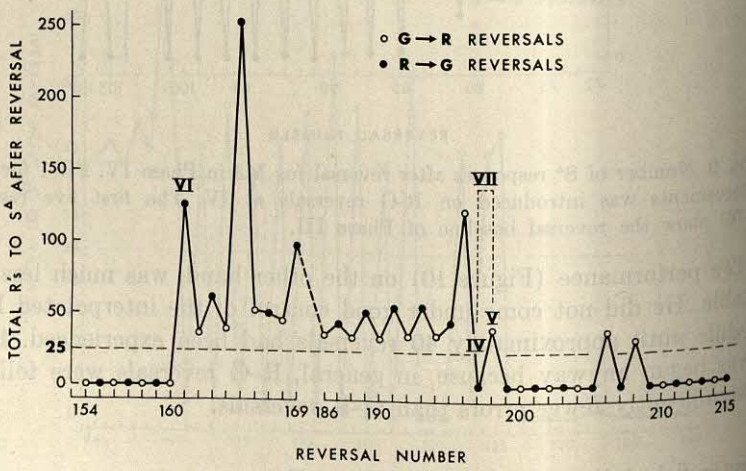


FIG. 11. Number of S^{Δ} responses for M1 in Phases VI and VII. At VI, the baseline or Phase III conditions were reinstated. At VII, Phases IV and V (FR 1 and the "count" of four reinforcements) were reintroduced on R-G and G-R reversals respectively.

At VI, the interpolated FR 1 and the regular programming of reversals were removed, re-establishing the baseline contingencies of Phase III. There was an immediate increase in errors followed by a recovery of the baseline (Reversals 186–196). While it appears that *S* was making somewhat fewer errors on G–R reversals than on R–G reversals in Phase VI, the number of responses to switch was highly similar to Phase III.

Reinstatement of the interpolated FR 1 on R–G reversals and reversals after a fixed number of reinforcements on G–R reversals almost immediately produced the near-perfect control achieved in Phases IV and V.

The performance of M2 in Phases VI and VII corroborates the effects shown in Figure 11. In Phase VI, M1's median errors were 46 (range 0–255) and M2's, 48 (range 0–196). In Phase VII, the medians were 0 (range 0–124) and 3 (range 0–52) respectively. M2, again, required more reversals than M1 in both Phase VI and Phase VII.

DISCUSSION

All of these retarded subjects demonstrated "learning set" (Harlow, 1950) by emitting fewer errors after repeated reversals. Although many reversals were required before responses to switch were sufficiently predictable to be used as baselines, sizeable decreases in errors occurred quite rapidly. The three subjects of Exp. 2 approximated asymptotic values after three reversals. However, errorless switching did not occur in either experiment until additional cues were added in Exp. 2.

There are no directly pertinent studies with which to compare the present results. Discrete trial studies, such as Ellis (1958), Girardeau (1959), and House and Zeaman (1959), were conducted with quite different procedures and all used FR 1 reinforcement exclusively. Long (1963) used chain FI FR and chain DRL FR schedules and was primarily interested in the specific stimuli which controlled such performances. Therefore, he exposed his subjects to only a few reversals.

Experiment 2 confirmed the finding of Exp. 1 that errors remain high when the ratio of responses to reinforcement is the only discriminative stimulus for switching. The median-FR 25 differences for M1, M2, and M3 were 40, 51, and 32 respectively which correspond closely to GM's excess responses to switch on FR 15 and FR 38. The similarity suggests that: (1) there is a lower limit to responses in excess of FR's; (2) order of exposure to schedules discussed earlier may not be an important contributing factor; and (3) continued exposure, alone, is not the key to reducing the discrepancy. Moreover, while "count" and interpolation of FR 1 resulted in errorless reversals when they were present, neither effected more accurate performance after they were removed. This has

been observed frequently in programmed instruction and has led to the use of "fading" procedures to transfer stimulus control from "prompts" to the stimulus desired (Holland, 1960). It seems that analogous procedures are necessary if one's aim is to achieve precise control by number of responses to reinforcement. Instructing subjects to count might alter the results significantly as was mentioned earlier, but that remains to be seen.

The tendency of subjects to switch "prematurely" is an indication that the schedule of reversals was a more salient stimulus than the ratio of responses to reinforcement. Certainly, the "count" procedure demonstrated that subjects readily come under the control of regular programming of reversals. One wonders, however, if this resulted from the low number of reinforcements (four) in contrast to the high number of responses (25). That is, would errorless switching be obtained with 25 reinforcements between reversals? The reversal studies using FR 1 do not lead one to predict a difference between reversals on FR 4 schedules of reinforcement and reversal after every fourth reinforcement. Comparisons such as these would be necessary to determine if the ratio of responses to reinforcement is inherently less likely to control behavior precisely. In any event, the "count" procedure demonstrated that an investigator must carefully control his schedule of reversals to obtain uncontaminated data.

Finally, the multiple baseline technique of Exp. 2 in which the discriminative properties of two events were compared simultaneously was not only more efficient but also eliminates sources of variation associated with comparisons separated in time as was the case in Exp. 1. That these subjects came under the control of the rather complex conditional discriminations required came as a surprise and testifies anew to the remarkable capacity of humans and other animals to respond to contingencies.

REFERENCES

Bijou, S. W., & Orlando, R. Rapid development of multiple schedule performances with retarded children. *Journal of the Experimental Analysis of Behavior,* 1961, **4,** 7–16.

Deterline, W. A. Operant discrimination reversals: comparative data. *Journal of the Experimental Analysis of Behavior,* 1960, **3,** 247–253.

Ellis, N. R. Object-quality discrimination learning sets in mental defectives. *Journal of Comparative and Physiological Psychology,* 1958, **51,** 79–81.

Ellis, N. R., Barnett, C. D., & Pryer, M. W. Operant behavior in mental defectives: Exploratory studies. *Journal of the Experimental Analysis of Behavior,* 1960, **3,** 63–69.

Ferster, C. B., & Skinner, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

GIRARDEAU, F. L. The formation of discrimination learning sets in mongoloid and normal children. *Journal of Comparative and Physiological Psychology,* 1959, **52,** 566–570.

HARLOW, H. F. Performance of catarrhine monkeys on a series of discrimination reversal problems. *Journal of Comparative and Physiological Psychology,* 1950, **43,** 231–239.

HOLLAND, J. G. Teaching machines: An application of principles from the laboratory. *Journal of the Experimental Analysis of Behavior,* 1960, **3,** 275–287.

HOUSE, B. J., AND ZEAMAN, D. Position discrimination and reversals in low-grade retardates. *Journal of Comparative and Physiological Psychology,* 1959, **52,** 564–565.

LONG, E. R. Chained and tandem scheduling with children. *Journal of the Experimental Analysis of Behavior,* 1963, **6,** 459–472.

MACKINTOSH, N. J. Selective attention in animal discrimination learning. *Psychological Bulletin,* 1965, **64,** 124–150.

OATLEY, K. G., BRYANT, P. E., & TINSON, C. Non-reinforcement and the emission of alternative responses by severely subnormal children. *Journal of Mental Deficiency Research,* 1965, **9,** 191–200.

ORLANDO, R., AND BIJOU, S. W. Single and multiple schedules of reinforcement in developmentally retarded children. *Journal of the Experimental Analysis of Behavior,* 1960, **3,** 339–348.

ORLANDO, R., BIJOU, S. W., TYLER, R. M., & MARSHALL, D. A. A laboratory for the experimental analysis of developmentally retarded children. *Psychological Reports,* 1960, **7,** 261–267.

PLENDERLEITH, M. Discrimination learning and discrimination reversal learning in normal and feebleminded children. *Journal of Genetic Psychology,* 1956, **88,** 107–112.

SHEPARD, W. O. Learning set in preschool children. *Journal of Comparative and Physiological Psychology,* 1957, **50,** 15–17.

SIDMAN, M. *Tactics of scientific research: Evaluating experimental data in psychology.* New York: Basic Books, 1960.

SPERLING, S. E. Reversal learning and resistance to extinction: a review of the rat literature. *Psychological Bulletin,* 1965, **63,** 281–297.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 514–531 (1969)

Effects of Increasing and Decreasing Reward Magnitude and Pre-experimental Persistence Level on Focal and Incidental Responses[1]

CHARLES Y. NAKAMURA
University of California, Los Angeles

Children rated as high persisters (HP) or low (LP) received increasing, decreasing, or random schedules of reward. Effects of reward, persistence, and sex factors on a focal lever moving response and an incidental latency response over rewarded and nonrewarded trials were investigated. Lever activity of LP girls directly reflected the schedules of reward magnitudes while HP girls showed high activity to low rewarded trials similar to findings in earlier work. Notably low lever activity of HP boys was attributed to their problem solving reaction to the task situation. The focal and incidental responses yielded different information about reward treatments under certain conditions. For LP boys, in particular, decrement in response speeds reflected the decreasing rewards while the concurrent lever activity either increased or remained stable. Similarities and differences between this and previous work were considered. Problems of comparing the results of this experiment and one that used analogous incentive schedules with animal *Ss* were also discussed.

This research investigated several effects of gradually increasing or decreasing schedules of reward over successive discrete trials. First was the differential effect of the schedules on two types of responses, lever activity as the focal response and latency as the incidental response. Second was the effect in the initial trials, late trials, trend over total rewarded trials, and in subsequent nonrewarded trials. Third was the comparison of the effects of changing reward schedules with previous findings that utilized constant magnitudes of relatively high or low reward.

A major interest was the use of these incentive procedures in the investigation of human behavior, particularly children. In a review of

[1]The research was supported in whole by Public Health Service Grant HD-00908 from the National Institute of Child Health and Human Development. The author is grateful to Roberta Evatt for running the *Ss*, Fred Shima for assisting in the data analysis, and Barry Godolphin for apparatus construction. The wholehearted cooperation of the teaching and administrative staffs of the University Elementary School is appreciated.

research on frustrative nonreward, Ryan and Watson (1968) observed that "The preponderance of research on frustrative nonreward has been restricted to discrete-trial instrumental appetitional performance of rats." They suggested the potential usefulness of extending the theoretical constructs and rigorous experimental procedures developed in the animal work to the study of children's behavior. The same recommendation is equally pertinent to the research on contrasted reward magnitudes and reward schedules since a great amount of systematic experimentation on animal behavior has been achieved (Pubols, 1964; Dunham, 1968), while work with children has been less systematic. Much of the work to date on reward effects with children has concerned the effectiveness of different kinds of rewards (Brackbill and Jack, 1958; Bijou and Sturges, 1959; Bisett and Rieber, 1966; Horowitz, 1962; Kennedy and Willcutt, 1964) and schedules of nonchanging magnitudes and percentages of reinforcement (Bruning, 1964; Chertkoff, 1968; Ryan and Voorhoeve, 1966). The effects of multiple reward schedules on continuous response rates on a free operant task are also well represented albeit much of the work was on retarded *Ss* (e.g., Bijou and Orlando, 1961).

Procedural differences in administering rewards have been systematically studied as seen in the reviews by Pubols (1964) and Dunham (1968). Pubols distinguished designs that used absolute (between *Ss*) and differential (within *Ss*) administrations of incentive magnitudes. Dunham distinguished among more complex contrasted conditions of reward: Simultaneous differential conditioning, successive differential conditioning, and successive nondifferential conditioning. These latter procedures take into account both cue and response contingencies and are applicable to either between or within subject designs. The primary design used in the present experiment was the differential method of Pubols or the successive nondifferential conditioning of Dunham since the same subjects experienced change in reward magnitude under the same cue-response conditions throughout the experiment.

The gradually changing reward schedules, in contrast to the sharply distinguishable, usually single, shifts used in most animal studies, were of interest since many experiences of humans are of this type. Some recent work with rats also suggests the importance of considering order of regularly changing rewards. Marx (1968) used regularly increasing or decreasing schedules of reward and found direct behavioral consequences of order during rewarded trials but level of nonreinforced performance was in a direction opposite of terminal level of reinforced responding. It was concluded that the findings support an interpretation of nonreinforced response rates in terms of order, rather than merely terminal value, of incentives during reinforcement. Using differential

reinforcement magnitudes, Ison and Glass (1968) found that increments and decrements in rewards did not have symmetrical behavioral consequences.

The selection of the two types of responses as the dependent variables followed from earlier work in which they were shown to provide different kinds of information about reward effects. Nakamura and Ellis (1964) found that lever activity could be considered an indicator of effort or performance level and response latency, time interval between the end of a trial and the press of a button to see the reward, an indicator of an attitudinal effect of reward. The utility of distinguishing the lever and the button press responses as focal and incidental was demonstrated by Nakamura and Krudis (1967). While the *S* focussed attention on the lever activity, the measure of latency was recorded without *S*'s knowledge. Thus, it provided a subtle index of reward treatments. In previous work (Nakamura and Ellis, 1964) children who were categorized as high persisters continued to put forth great effort under low rewards as indicated by lever activity but their attitudes toward the task were changing as reflected by the gradually declining response speeds to press the button that revealed their rewards.

METHOD

Subjects

The *S*s were 24 boys and 24 girls, of age range 5 years 6 months to 6 years 11 months with mean age 6 years 6 months, enrolled in the UCLA University Elementary School. The mean intelligence was well above average. The large majority of *S*s were from families above average in socio-economic and educational levels.

Apparatus

The main apparatus consisted of a modified version of an automatic marble reward display and dispenser unit, a lever movement task, and electrical control unit that were described in a previous experiment (Nakamura and Ellis, 1964). Eight separate marble dispensers were side by side on a table. Each dispenser was a box 5½-inches wide, 17-inches high and 4-inches deep. A clear plastic ¾-inch tube, 7¾-inches high was mounted above a marble receptacle at the base of the box. On display within each tube were different numbers of 9/16-inch diameter opaque green marbles that indicated the number available per reward. Straddling each tube was a pair of ½-inch diameter green reflectors which would light up as the marbles from a given tube were simultaneously ejected into the tray below when *S* pushed a button to receive his reward.

The emptied tube was automatically refilled from a magazine concealed in the top part of the dispenser box.

The lever movement task consisted of a 7-inch diameter metal, circular plate set in the surface of a 9-inch cubical wooden box. The plate had eight 3-inch long slots that radiated symmetrically from a 1-inch diameter opening through which protruded a lever. The lever was anchored at the base of the box and extended 5 inches above the plate. It was held at the central resting position by coiled springs. Within the box was a spring loaded switch at each slot. Entrance of the lever approximately 1½ inches into a slot contacted the corresponding switch. Each contact was recorded on an impulse counter in the control unit. At each upper corner of the box's surface above the plate was a light that signalled *S* when to start and to stop lever activity. Located centrally and near the top of the front side of the box was the button by which *S* actuated the reward dispenser. On the table was a 7-inch cubical marble storage box with a hole in top into which *S* dropped the marbles. All of these boxes in the apparatus were painted gray.

Procedure

Teachers' ratings of the *Ss*' general predisposition to persist were obtained on a single 9-point scale from low persistence (LP) to high persistence (HP). The ratings were made two months after the beginning of the semester so the teachers were well acquainted with the children. Persistence as rated was closely related to variables such as need-achievement and history of successful experiences. Information of reliability and content validity of the scale were reported earlier (Nakamura and Ellis, 1964). On a median cut, *Ss* rated 5 or lower and those rated 6 and higher were categorized as LP and HP, respectively. From the total available population, 12 LP and 12 HP girls and boys were selected for the experiment so that subgroup means for boys and for girls within each level were similar. Within each of these four groups, four *Ss* were assigned to each of three reward treatments. This yielded a factorial design with two levels of rated persistence, three schedules of reward, and two sexes.

The eight dispenser boxes comprised a display panel of from 0 to 7 marbles in the tubes from left to right. The *S* received one of these magnitudes of reward per trial over 15 trials on one of three schedules on which the distribution of magnitudes differed but the mean number of marbles received were the same. The schedules were Increasing rewards (IR): 3, 2, 1, 1, 2, 3, 3, 4, 6, 5, 5, 7, 5, 6, 7; Decreasing rewards (DR): 5, 6, 7, 7, 6, 5, 5, 4, 2, 3, 3, 1, 3, 2, 1; and Randon rewards (RR): 3, 7, 2, 6, 1, 5, 5, 7, 2, 5, 1, 5, 3, 6, 4.

The *S* was seated facing the reward display with the surface of the lever box about lap height and the lever handle directly in front. The initial instructions were given in the experimental room, after which the *E* went into and remained in the adjoining room separated by a one-way vision window and gave subsequent instructions over two-way intercom. The initial instructions were:

"Show me the tube with the most marbles in it. How many are in the tube? Now, show me which one has the smallest number of marbles. You see there are two lights here (surface of lever box). When these lights go out you can begin pushing this handle into these slots like this (demonstrated). You can push it into any slot you want to, as often as you want, and as fast as you want to. Each slot gives you points, but some slots give you more points than others. I can't tell you which ones give you the most points. You will have to figure that out for yourself. Now you try it (practice). You keep pushing the handle as long as these lights are out. When the lights come back on again, you stop pushing the handle. You then push this red button (dispenser button). Go ahead and try it now. This will make the machine give you your marbles. Now when the marbles fall down into these trays, you pick them up and put them in the gray box over there. Let them drop one by one through the hole in the top. Be sure and watch the lights closely for they will tell you when to start and to stop pushing the handle. Remember, you start pushing the handle when the lights go out. O. K. we are going to practice this now in make believe (*E* goes through the whole procedure with *S* step by step). Do you have any question? O. K. I'm going out of the room now. You can talk to me through this box (intercom). I can talk to you through it too. If you have any questions ask them right away. Watch the lights and begin playing when they go out."

The first two trials were practice and the *S* received four marbles on each. Then the 15 regular rewarded trials were given in the assigned schedule. Thus, the rewards were not contingent upon actual performance. Each trial consisted of 8 seconds of free lever activity. The number of strokes into the slots was recorded on the impulse counter. The latency between the onset of the signal lights that indicated the end of a trial and pressing of the reward dispenser button was recorded by *E* to the nearest .10 second from a Hunter Klockounter. An extinction (nonreward) phase began at the end of the 15 rewarded trials with the following instructions:

"Now that you have had some practice I'm going to set the machine so that it will be harder for you to get points. We can stop anytime you want to, but I would like to see how well you can do when I make it harder. Are you ready?"

After every third trial one of the following statements was made by *E:* "Would you like to try it again or would you rather stop now:" or "You can keep on playing or stop whenever you want."

In the extinction procedure, when *S* pressed the reward dispenser button, the lights went on in the zero marble dispenser and no marbles were dispensed. In addition to the number of lever strokes and latency on each trial an extinction score, number of trials to extinction, was obtained on each *S*. At the end of the period, when *S* had either requested to terminate or had gone the limit of 15 nonrewarded trials, two extra trials were given with these instructions: "The last game was very hard —everyone has trouble getting marbles there. But there is one more game I would like you to try. It is different from the last one. We'll play this game just two times." High rewards were given so that *S* would leave with a feeling of success. Then the *E* had *S* help remove the marbles from the storage box to reload the dispenser magazine. This was accompanied by complimentary remarks from *E* on the large number of marbles *S* had earned and that *S* had done very well. Almost all *S*s seemed to be fascinated by the dispenser mechanisms when the boxes were opened and they enjoyed reloading them.

RESULTS

The analysis of variance applied to the lever stroke scores yielded several significant main effects and interactions as shown in Table 1. All results of $p < .05$ will be reported as significant. The entries were mean lever strokes per block of three trials. The group means for boys and girls separately and combined are plotted in Fig. 1. The nature of the significant Reward effect ($F = 4.97$) can be best seen in the lower panel. The Decreasing reward (DR) treatment has the lowest lever strokes and the Randon reward (RR) has the highest with the Increasing (IR) group between them. The middle and upper panels taken together show the source of the significant Sex $\times$ Persistence interaction ($F = 16.55$). Lever strokes of Low Persistence (LP) boys were greater than for High Persistence (HP) while the converse obtained for the girls.

The significant Blocks effect ($F = 16.81$) reflects the gradual overall increase in means apparent in the lower panel. The sources of the Blocks $\times$ Reward $\times$ Persistence interaction ($F = 2.18$) and the four-way interaction ($F = 3.70$) can be ascertained in the panels of Fig. 1. Their implications will be left to the discussion section.

Lever strokes in the early trials were assessed by an analysis of variance applied to the mean of Blocks 1 and 2 combined. A significant Reward main effect ($F = 4.55$, $df = 2,36$) and a significant Sex $\times$ Persistence interaction ($F = 9.26$, $df = 1,36$) were obtained. The sources

TABLE 1
ANALYSIS OF VARIANCE OF LEVER STROKES

Source	*df*	MS	*F*
Between			
Reward (R)	2	394.50	4.97*
Sex (S)	1	11.00	.14
Persistence (P)	1	119.00	1.50
R × S	2	52.50	.66
R × P	2	16.00	.20
S × P	1	1312.00	16.55**
R × S × P	2	87.50	1.10
Error (b)	36	79.28	
Within			
Blocks (B)	4	109.75	16.81**
B × R	8	4.88	.75
B × S	4	15.25	2.34
B × P	4	9.75	1.49
B × R × S	8	9.63	1.47
B × R × P	8	14.25	2.18*
B × S × P	4	13.00	1.99
B × R × S × P	8	24.13	3.70**
Error (w)	144	6.53	

* $p < .05$.
** $p < .01$.

of these effects were similar to those observed above in connection with the corresponding interactions in the analysis over total trials. Thus, it appeared that the relative levels of activity for the groups were set in the early trials and were not appreciably changed over the later trials. This was further confirmed by the analysis of the terminal performance scores in Block 5 since the Reward effect ($F = 3.42$, $df = 2,36$) and the Sex × Persistence interaction ($F = 14.51$, $df = 1,36$) were again significant.

The group means for lever strokes in the first five nonrewarded trials are plotted to the right of the broken line in Fig. 1. The entries were the mean score of from one to five trials for each *S*. The number of trials varied because it was possible for an *S* to extinguish anytime after the first nonrewarded trial. Only the Sex × Persistence interaction ($F = 4.18$, $df = 1,36$) was significant. An analysis of variance applied to the difference scores between the Block 5 and the nonrewarded trials yielded no significant results.

The main findings in the analysis of lever strokes were: (1) The order of magnitude of strokes from greater to least was RR, IR, and DR treatment; (2) activity level was greater in the LP than the HP boys while

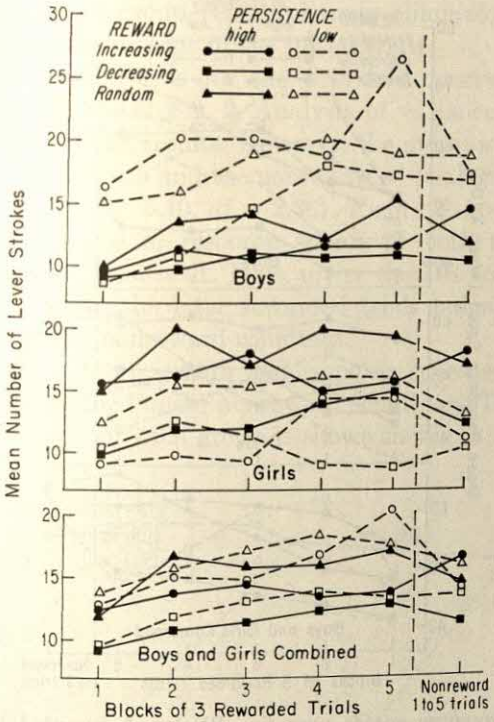


FIG. 1. Mean number of lever strokes over five Blocks of rewarded trials and for 1–5 nonrewarded trials.

the reverse order was found for girls, and these relative positions were maintained throughout the rewarded trials and into the first five nonrewarded trials; and (3) complex interactions involving all factors were obtained.

The overall analysis of variance of the response speed scores (reciprocals of the response times between end of each lever activity trial and press of reward dispenser button) yielded a significant Blocks main effect ($F = 4.04$, $df = 4,144$) and a significant Blocks $\times$ Reward $\times$ Persistence interaction ($F = 2.74$, $df = 8,144$). The mean responses in blocks of three trials are plotted in Fig. 2 for boys and girls separately and combined. The curves in the lower panel indicate that an overall increase in means over blocks was the source of the main Blocks effect. Also apparent is the source of the three-way interaction. This was largely the function of the more curvilinear trend of the LP *Ss* in the DR treatment (curve with open squares) compared to the other curves in general and to the HP-DR curve in particular. A trend test of the quadratic component (Grant, 1956) supported this observation. The main Reward effect was

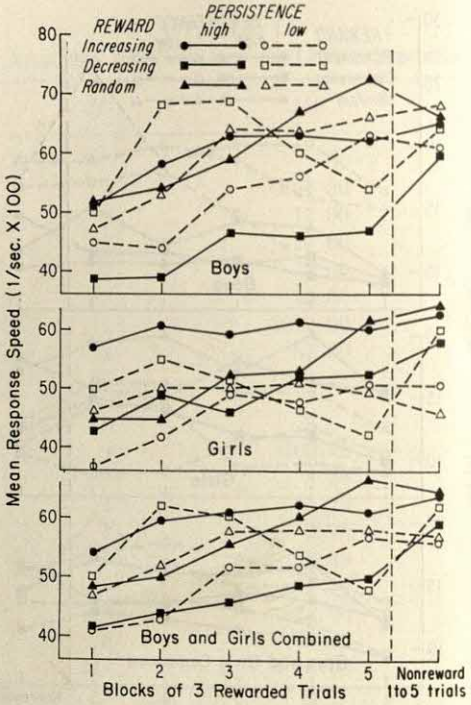


FIG. 2. Mean response speeds over five Blocks of rewarded trials and for 1-5 nonrewarded trials.

significant ($F = 3.45$, $df = 2,36$), but of particular importance with regard to the above inference was the significant Reward $\times$ Persistence interaction ($F = 6.11$, $df = 2,36$). The trend of initial incline followed by continuous decline is apparent in both LP-DR boys and LP-DR girls (Fig. 2, upper and middle panels) and are the only curves that show a clear decline over the later blocks of trials.

The results of the analyses of the response speeds in the early trials and in the terminal trials were also quite consistent with the above impressions of trends. Whereas the analysis of Blocks 1 and 2 combined scores yielded a significant Reward $\times$ Persistence interaction ($F = 5.25$, $df = 2,36$), it was clearly not significant ($F = .25$, $df = 2,36$) in the corresponding analysis of the Block 5 scores. It can be seen in the lower panel of Fig. 2 that the locus of the interaction in the early trials was that mean response speed of the LP-DR group was greater than for the HP-DR group while the reverse obtained for the LP-IR and the HP-IR groups. However, by Block 5 the LP-DR mean and the distance between the HP-IR and LP-IR groups had closed considerably owing to a gradual

increase in the LP-IR group. These changes eliminated the Reward × Persistence interaction present in the initial trials.

The mean response speeds for the first five nonrewarded trials are plotted at the right side of Fig. 2. Analysis of variance applied to these scores gave no significant results. However, the analysis of the difference scores between the Block 5 and the nonrewarded trials produced a significant Reward effect ($F = 5.10$, $df = 2,36$). Figure 2 shows that the main source was the increase in response speeds of both the HP and LP groups under the DR treatment. Thus, under the DR treatment response speed was depressed in the later rewarded trials but sharply increased with the change to the nonreward condition.

In addition to the lever strokes and response speed scores, a measure of extinction was obtained under nonreward conditions. The mean number of trials to extinction for each group is shown in Fig. 3. The scores were

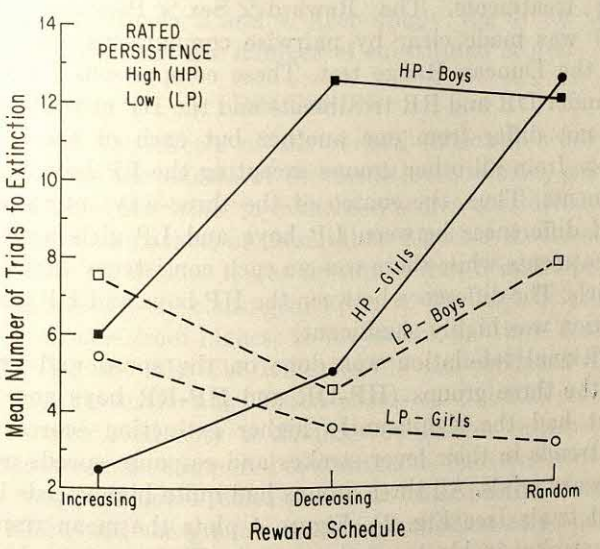


FIG. 3. Mean number of trials to extinction for high persistence and low persistence boys and girls subsequent to three schedules of rewarded trials.

submitted to a square root transformation. The summary of the analysis of variance of these scores is given in Table 2. The main Sex effect ($F = 5.04$) reflects the higher scores of boys over girls. The source of the main Persistence effect ($F = 4.44$) was the higher scores of HP compared to LP *Ss*. Figure 3 shows that the source of the Reward × Persistence interaction ($F = 3.39$) was the regularly increasing scores of the HP *Ss* from IR through DR to RR treatment groups in contrast to the

TABLE 2
ANALYSIS OF VARIANCE OF TRIALS TO EXTINCTION

Source	*df*	MS	*F*
Reward (R)	2	2.72	2.83
Sex (S)	1	4.84	5.04*
Persistence (P)	1	4.26	4.44*
R × S	2	.25	.26
R × P	2	3.25	3.39*
S × P	1	.03	—
R × S × P	2	4.61	4.80*
Error	36	.96	

* $p < .05$.

shallow V-shaped distribution for scores for the LP *Ss* over the corresponding treatments. The Reward × Sex × Persistence interaction ($F = 4.80$) was made clear by pairwise comparisons among the group means by the Duncan Range test. These comparisons showed that the HP boys under DR and RR treatments and the HP girls under RR treatment did not differ from one another but each of the three differed significantly from all other groups excepting the LP boys under IR and RR treatments. Thus, the source of the three-way interaction was the absence of differences between LP boys and LP girls across the three reward treatments while there was no such consistency between HP boys and HP girls. The difference between the HP boys and HP girls under the DR condition was highly significant.

An additional tabulation was done on the speed and lever activity scores of the three groups (HP-DR and HP-RR boys and the HP-RR girls) that had the significantly higher extinction scores in order to ascertain trends in their lever strokes and response speeds over continuing nonreward trials. All three groups had quite high speeds in the initial nonreward trials (see Fig. 2). Figure 4 plots the mean response speeds and lever strokes in blocks of three trials. The apparent deceleration of response speed over Blocks was confirmed by a test of the trend ($F = 6.46$, $df = 1,40$). In contrast, lever strokes were maintained at about the same level throughout. This response speed curve tended to parallel that of the LP-DR curve (Fig. 2, lower panel) over the last four blocks of rewarded trials. These results were only tentative because of some missing scores in the later blocks of the nonrewarded trials. The mean scores of a few *Ss* in Blocks 3 and 4 in Fig. 4 were based on only two trials or one instead of three since they had fewer than 12 trials to extinction. Furthermore, the Block 4 group mean value was substituted for the Block 4 scores of two *Ss* who had only seven and eight extinction trials, respec-

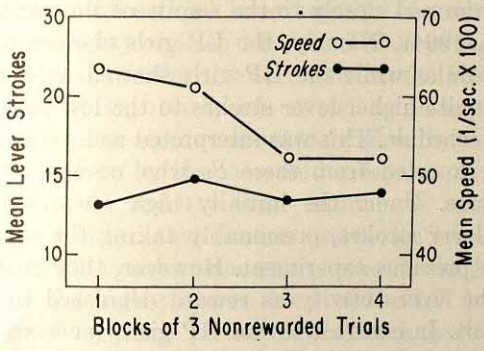


FIG. 4. Mean lever strokes and mean response speeds over four Blocks of nonrewarded trials for 12 *S*s from the three groups that were most resistant to extinction.

tively. For one *S* who had only six trials, the block means had to be substituted in both Blocks 3 and 4. Accordingly, the *df* for the above *F* test was reduced by four, the number of substituted scores.

DISCUSSION

Two aspects of the effects of the reward schedules were of central interest. They were the effects of the focal lever strokes, the incidental response speeds, and the trials to extinction scores considered separately, and the various combinations of these taken together. Lever activity was the focal response for *S* since it was presumed that *S* would perceive that the rewards received were contingent upon it. Throughout the trials the order of lever strokes from highest to lowest for the three reward schedules was RR, IR, and DR. The RR schedule also produced the greatest number of trials to extinction. These effects of the RR schedule were interesting in their similarity to findings of greater response produced by partial reinforcement compared to continuous reinforcement schedules. The higher response level corresponds to the partial reinforcement acquisition effect, and the greater trials to extinction corresponds to the partial reinforcement extinction effect typically obtained in the studies of frustration (Amsel, 1962; MacKinnon and Amsel, 1964; Ryan and Watson, 1968). Apparently, the symmetrical proportion of high and low rewards in the RR schedule had an effect similar to that of the 50% partial reinforcement schedule obtained by Ryan and Voorhoeve (1966). They found that a 50% schedule yielded the greatest speed on a lever movement response while 100% reinforcement gave the lowest speed.

Boys and girls performed differently on lever strokes as a function of persistence level. The LP boys consistently made more lever strokes than HP boys while the converse was generally true for the girls. The interaction of reward and persistence over blocks of trials of the girls

(see Fig. 1) conformed closely to the results of the earlier work (Nakamura and Ellis, 1964). Namely, the LP girls showed a direct effect of the reward schedules while the HP girls showed a contrast effect. The HP girls made quite higher lever strokes to the low reward in the initial trials of the IR schedule. This was interpreted as a reaction to frustration that would be expected from these *Ss* who were presumed to have a history of success. Under the initially high rewards in DR, the HP girls gave low lever strokes, presumably taking the success for granted as found in the previous experiment. However, they reacted with an increasing level of lever activity as reward decreased to low magnitudes in the later trials. In contrast to the HP girls, lever strokes of LP girls increased with the increasing rewards in IR and decreased with the decreasing rewards under DR. The differences in lever activity between boys and girls on interactions of reward schedules and persistence levels over the blocks of trials were the main sources of the significant triple and four-way interactions (see Table 1) noted in the results section. The performance of the girls was consistent with the results and the rationale of the earlier work but that of boys was not. In previous work it was assumed that the receipt of low rewards would be frustrating because it implied failure. Accordingly, the HP girls responded with high activity. However, the HP boys unexpectedly gave a low level of lever activity. One plausible explanation is in terms of a problem solving oriented reaction to the low rewards. There is evidence that boys respond this way more than do girls (e.g., Crandall and Rabson, 1960; Nakamura and Ellis, 1964). There is also clear evidence that boys are superior to girls on problem solving tasks of the types generally used in experiments (Gallagher, 1964, p. 370; Witkin, Dyk, Faterson, Goodenough, and Karp, 1962). Furthermore, children of high persistence were found to be more problem solution oriented than low persisters (Nakamura and Lowenkron, 1964). These findings, together with the observational data obtained on the *Ss*, were consistent with the above interpretation of the present results. Throughout the experiment *E* regularly recorded the more salient behaviors of each *S*. The tabulation of these revealed that the HP boys as a group were distinguished by appearing to be very concentrating, thoughtful, planning and looking for slots that gave the most points. Of the HP boys, 67% were characterized by two or more of these terms in contrast to 33% of HP girls and 21% of LP *Ss*. The HP boys could, accordingly, be described as being very problem solution oriented. For example, many of them were noted to have frequently paused for a few seconds holding the lever motionless; some subsequently making lever strokes into apparently preconceived combinations of successive slots. This strategy was reflected in the relatively low lever strokes of most

HP boys in contrast to the rapid lever activity of most *Ss* in the other groups. An alternative explanation that the HP boys were disinterested might be entertained, particularly with regard to the HP-DR boys since they had low scores on response speeds as well as on lever strokes. However, evidence that the low scores were associated with involvement rather than disinterest was given by the increase in response speeds in the nonrewarded trials (Fig. 2) and the large number of trials to extinction (Fig. 3). It may be recalled that these nonrewarded trials were preceded by instructions that the problem was going to be harder than before.

This still leaves the high response level of the LP boys unexplained since the corresponding group in the earlier work gave a low level of activity to the low reward much as the LP girls did in both experiments. In retrospect, the most probable source of this discrepancy was in the rated persistence level of the LP boys in this experiment. Within each experiment, HP and LP groups were established by division at the median persistence score. A comparison of the scores of subgroups revealed that the mean of LP boys in this experiment was only slightly and nonsignificantly smaller than that of the HP *Ss* but clearly larger than that of the LP *Ss* in the previous experiment. This occurred because the overall mean persistence was higher in the present University Elementary School boys than in the previous parochial school boys, thus, resulting in the relatively higher mean persistence score in the LP group of the present *Ss*. A difference between LP groups did not occur in the girls because the overall means were similar in the two schools. This fact of the lack of difference in persistence between HP and LP boys in the two experiments gives a plausible explanation for the similarity of their performances. This reasoning is given very tentatively but it gains some credence from examination of the response speed curves for these two groups since they, as well as the lever activity curves, were quite similar. In fact, the most salient result in the analyses of response speeds was the curvilinear trend in the LP-DR groups. The initial increase with later decline was particularly prominent in the curve for the boys. In past work (Nakamura and Boroczi, 1965; Nakamura and Ellis, 1964) the occurrence of such a decline in response speed was accompanied by a concurrent increase in performance on the focal response under low reward treatments. Thus, while performance on the focal lever response changed in a direction opposite to the change in rewards, the decreasing magnitudes of reward were reflected in the more subtle response speed measure that was considered to be an attitudinal aspect of the reward effect. At least for LP boys, these two response measures gave different information about the reward treatment as in the earlier work. There is no clear explanation for the absence of this in the HP boys and HP girls. This

could, in part, be due to the changing reward schedules here compared to the constant magnitudes used in the previous work. Of additional interest in this regard was that, following the increase in speed at the beginning of the nonrewarded trials, the HP *Ss* with high resistance to extinction showed a decline in response speed while the concurrent level of lever strokes remained stable (Fig. 4). This indicated that although the attitude change (response speed) that was expected to be associated with the DR schedule in the HP *Ss* did not take place during the rewarded trials, as was found for the LP-DR *Ss*, it was eventually manifested in the subsequent trials under nonreward. This was consistent with the rationale set forth in the earlier report by Nakamura and Krudis (1967) that the focal activity could be expected to increase or remain stable for long periods even under low rewards in highly persistent, achievement oriented *Ss* while a negative attitudinal effect of the low reward would be manifested in the incidental response speed measure.

Recommendations have frequently been made promoting the extension of the rigorous experimental designs and procedures developed in animal studies to the investigation of children's behaviors (McCandless and Spiker, 1956; Nakamura and Ellis, 1964; Terrell, 1958). Such an extension of incentive treatment procedures was one of the interests in this study. However, the results of work using similar methodologies on animals and children may not be comparable or compatible in their theoretical implications. One reason for this is that the pre-experimental histories of animals, e.g., the naive rat, are typically more uniform and also more controllable than in children, hence, individual differences on critical variables are smaller. Accordingly, procedures used with naive animals should, when applied to children, take into consideration the histories of the child *Ss*. This immediately complicates the experimental design and often requires alterations in the predictions based on the conceptual frameworks developed on animals. This problem was illustrated when the attempt was made to compare predictions and results of this experiment with one on rats (Marx, 1968) that used fairly analogous reward procedures of increasing and decreasing schedules of reinforcement, and the use of an activity measure of performance during both rewarded and nonrewarded treatments. There were to be sure some obvious procedural differences. An interval schedule with free operant responding, training and test sessions spaced over many days, and a gustatory reinforcer were used with the rats. The present experiment used massed discrete trials in a single session and the magnitudes of rewards were mediated through the visual modality. The several training sessions required for the rats in order to establish the conditions for the critical effects to appear were unnecessary for children because the latter can be verbally instructed about procedures.

On rewarded trials Marx (1968, cf. Fig. 1) found a direct effect of reward schedules; an increasing number of bar presses over trials for the group receiving an increasing order of reinforcement and decreasing number of presses for the group receiving decreasing amounts of sucrose reinforcement. With children, similar results were predicted for LP *Ss* but not for HP *Ss*. These differential predictions were generally supported. On nonrewarded trials, while both experiments were consistent in proposing, as stated by Marx, the usefulness of considering response strength in terms of order rather than simply terminal value of incentives during reinforcement, the expected results were not altogether the same. Marx expected and found that nonreinforced response rate would be higher with increasing than decreasing order of rewards. In the present experiment an interaction of reward schedule and persistence level was predicted. The predicted interaction was obtained for resistance to extinction, although it was not found on the measure of lever strokes. As expected, the results for the LP *Ss* were in the same direction (LP-IR > LP-DR) as Marx's results while the reverse held for the HP *Ss* (HP-DR > HP-IR, see Fig. 3).

These comparisons indicated that variability in predispositions such as persistence level that exist in children complicate and set limitations to making direct comparisons with most of the animal studies on incentive effects. In this regard, a combination of research on early experiences and the work on incentive magnitudes in rats could be fruitful. Other well recognized factors that complicate comparisons between experiments using animal and children *Ss* are social variables such as the tendency for performance to be influenced by degree of concern of how others perceive and respond to oneself. In this realm, the use of gustatory reinforcers can be expected to have relatively limited value for the understanding of much of children's behaviors. On the other hand, reinforcers mediated through other modalities have the problem of very wide individual differences in relative incentive values for any given absolute quantity or quality of the reinforcer. For example, an amount of money, marbles or trinkets, or a statement of good, right, or wrong, or a smile or a frown have different incentive values for different children. It would appear that, in general, extension of methodology for animal experiments to children can contribute useful experimental precision but the conceptual frameworks and predictions of behavior must be generalized with reservation.

REFERENCES

AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and theoretical extension. *Psychological Review*, 1962, **69**, 306–328.

BIJOU, S. W., AND ORLANDO, R. Rapid development of multiple-schedule perfor-

mance with retarded children. *Journal of the Experimental Analysis of Behavior,* 1961, **4**, 7–16.

Bijou, S. W., and Sturges, P. T. Positive reinforcers for experimental studies with children—consumables and manipulatables. *Child Development,* 1959, **30**, 151–170.

Bisett, B. M., and Rieber, M. The effects of age and incentive value on discrimination learning. *Journal of Experimental Child Psychology,* 1966, **3**, 199–206.

Brackbill, Y., and Jack, D. Discrimination learning in children as a function of reinforcement value. *Child Development,* 1958, **29**, 185–190.

Bruning, J. L. The effects of magnitude of reward and percentage of reinforcement on a lever movement response. *Child Development,* 1964, **35**, 281–285.

Chertkoff, J. M. Response speed under voluntary and involuntary conditions as a function of percentage of reinforcement. *Child Development,* 1968, **39**, 167–175.

Crandall, V. J., and Rabson, A. Children's repetition choices in an intellectual achievement situation following success and failure. *Journal of Genetic Psychology,* 1960, **97**, 161–168.

Dunham, P. J. Contrasted conditions of reinforcement: A selective critique. *Psychological Bulletin,* 1968, **69**, 295–315.

Gallagher, J. J. Productive thinking. In M. L. Hoffman and L. W. Hoffman (Eds.), *Child development research.* Vol. 1, New York: Russell Sage Foundation, 1964. Pp. 349–381.

Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychological Bulletin,* 1956, **53**, 141–154.

Horowitz, F. D. Incentive value of social stimuli for preschool children. *Child Development,* 1962, **33**, 111–116.

Ison, J. R., and Glass, D. H. Long term consequences of differential reinforcement magnitudes. *Journal of Comparative and Physiological Psychology,* 1968, **65**, 524–525.

Kennedy, W. A., and Willcutt, H. C. Praise and blame as incentives. *Psychological Bulletin,* 1964, **62**, 323–332.

MacKinnon, J. R., and Amsel, A. Magnitude of the frustration effect as a function of degree of confinement and detention in the frustrating situation. *Journal of Experimental Psychology,* 1964, **67**, 468–474.

Marx, M. H. Nonreinforced responding as a function of the direction of a prior ordered incentive shift: A replication with fixed-interval reinforcement schedule. *Journal of Experimental Psychology,* 1968, **77**, 159–160.

McCandless, B. R., and Spiker, C. C. Experimental research in child psychology. *Child Development,* 1956, **27**, 78–80.

Nakamura, C. Y., and Boroczi, G. Effect of relative incentive value on persistence and response speed. *Child Development,* 1965, **36**, 547–557.

Nakamura, C. Y., and Ellis, F. F. Methodological study of the effects of relative reward magnitude on performance. *Child Development,* 1964, **35**, 595–610.

Nakamura, C. Y., and Krudis, B. R. Evaluation of a response speed measure of incentive value of reward. *Journal of Experimental Psychology,* 1967, **74**, 44–49.

Nakamura, C. Y., and Lowenkron, B. Z. Incentive magnitude, task orientation, and persistence. *Child Development,* 1964, **35**, 611–621.

Pubols, B. H. Incentive magnitude, learning, and performance in animals. *Psychological Bulletin,* 1960, **57**, 89–115.

Ryan, T. J., and Voorhoeve, A. C. A parametric investigation of reinforcement

schedule and sex of *S* as related to acquisition and extinction of an instrumental response. *Journal of Experimental Child Psychology,* 1966, **4,** 189–197.

RYAN, T. J., AND WATSON, P. Frustrative nonreward theory applied to children's behavior. *Psychological Bulletin,* 1968, **69,** 111–125.

TERRELL, G., JR. The need for simplicity in research in child psychology. *Child Development,* 1958, **29,** 303–310.

WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., AND KARP, S. A. *Psychological differentiation.* New York: Wiley, 1962.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 532–540 (1969)

Effects of Time-Compressed Speech Signals on Children's Identification Accuracy and Latency Measures

THOMAS H. SHRINER AND ROBERT L. SPRAGUE[1]
University of Illinois

Identification accuracy and latency of responses were studied under various rates of time-compressed speech signals. *S*s were 51 elementary school children, 10 kindergarteners, 16 first-graders, 14 second-graders, and 11 third-graders. Each *S* was instructed to respond to one of a set of three line drawings selected from the Picture Identification Children's Standardized Index when hearing a speech signal. The speech signals, 50 in number for each treatment plus 24 appropriate adaptation trials, were presented under normal speech and the three experimental conditions (30, 50, and 70% time-compression). Each *S* received only one treatment, exactly 14 days after the normal run. Mean identification accuracy, based on corrected percentage scores, significantly decreased as a function of time-compression. Graphically, very little decline appeared before the 50% condition. Mean latencies for incorrect as well as for incorrect and correct responses combined were significantly lower for the 50% treatment than for the 30 and 70% treatments, and the means for the 30% condition were lower than those for the 70% condition. The incorrect mean latency scores also were significantly different from the correct mean latency scores.

Increasing emphasis is being given to rate-controlled recordings by both educators and experimenters. One marked advantage of rate-controlled recordings (time-compressed speech) is that recorded materials can be played back in less than the original time without essential alteration of the other dimensions of the signal. In a sample of speech, the time base is unselectively and uniformly compressed throughout. The essentials of the phonetic units, vocal pitch, stress, etc., are preserved. The amount of compression, which may conveniently be thought of as a percentage of the original time, is variable and a given message may be presented to listeners at a variety of rates. Practically, very little is known, however, about the limitations and/or applications of time-compressed speech signals to normal elementary school children or to children who somehow fail to meet the criteria of normalcy. Moreover,

[1]This investigation was supported by Public Health Service Research Grant NB 0-7346 from the National Institute of Mental Health and Grant G-3-6-062063-1559 from the Office of Education.

the nature of the relationship between rate and accuracy of immediate recall is highly critical for any theory of short-term memory (Aaronson and Markowitz, 1967; Aaronson, 1967).

In conventional intelligibility tests, the response measure is the percentage of test words correctly identified. Hecker, Stevens, and Williams (1966) suggest "the inclusion of reaction-time (latency) measurements in multiple-choice intelligibility tests should increase the sensitivity of such tests" (p. 1189). In a brief study with five adults, they reported an overall trend of an increase in reaction time with an accompanying decrease in intelligibility. The question arises whether measures of latency of response when used with the conventional percentage of correct response measures are a more sensitive index for assessing children's abilities to discriminate speech signals. The purpose of this study is to investigate the accuracy and latency of children's identifications of speech signals presented under three different rates of time-compression.

METHOD

Subjects

The *S*s were 51 elementary school children, 10 kindergarteners, 16 first-graders, 14 second-graders, and 11 third-graders.[2] No child had an educationally significant hearing loss as reported by school personnel.

Stimuli

The stimuli used were the Picture Identification Test for Children—A Standardized Index (PICSI) which was developed by Seidel (1963) to test consonant discriminative abilities in children. The PICSI consists of 50 word-triads in the form of "cat-bat-hat" with associated line drawings, white outline on black background, for each member of the triad. Eight of the triads have a common initial consonant (cave-cane-cake, for example) and 35 have a common ending consonant (gate-plate-skate). The PICSI also has eight triadic practice items which are not included in the test proper. Slides were made for each of the 50 word-triads and 24 slides (three for each) were made of the eight practice items.

Speaker and Apparatus

The speaker was a female graduate student who spoke general American speech. She was seated in a sound-treated chamber and read

[2] We thank Mr. Eugene Rankin, Principal, and the teachers of Lincoln School, Urbana, Illinois, and also thank Mr. George Montgomery and Mr. Daniel Beasley, Research Assistants, for their assistance. The authors also thank Morton Weir who critically read the first draft of the manuscript.

List 1 of the PICSI which consisted of the 50 test items and 8 practice items. Each item was preceded by the carrier phrase "Show me—." After practice and visually monitoring the speech signals on a VU meter by both the speaker and the *E*s, recordings were made on a Magnecord M-90 tape recorder (frequency response ±3 dB, 40–11,000 Hz at 15 ips) which was located in an associated control room. An Altec M-11 condenser microphone was used. The speech signals were spoken at a conversational effort and pitch level, with natural inflection. (The estimated fundamental frequency of phonation was approximated from a large drum sonogram—Kay Electric Sonograph 6061A—to be about 250 Hz.) The recordings were repeated until judged acceptable by both the *E*s and the speaker. The average duration of the carrier phrase plus the PICSI word was 1.19 seconds. The duration was determined by playing the tape through a Sound Level Apparatus Company HPL-E high speed level recorder.

The master tape was then used to make the experimental tapes. This was accomplished by time-compressing the master tape three times, once for each condition (30, 50, 70%). Time compression was accomplished by using the Fairbanks, Everitt, and Jaeger (1954) Speech Compressor utilizing a 20-millisecond interval of discard. The compression unit effectively records and discards alternate segments of a signal, rejoining the recorded segments into a continuous output signal. For a given interval of speech (I_a), the relationship between the recorded (I_r) is $I_a = I_r + I_d$, and the percentage time compression of interval I_a is

$$TC\% = \frac{I_d}{I_r + I_d} \times 100$$

Each tape was high pass filtered (300 Hz, 36 dB per octave) to remove low frequency noise generated by the compression system. The signal-to-noise ratios for the 30, 50, and 70% conditions were 40, 40, and 39 dB, respectively.

For each of the time-compressed conditions, a single sequence of word-triad items was randomly determined. The position of the correct choice within each triad of drawings (left, center, right) was counterbalanced. The eight practice items were presented three times for a total of 24 adaptation trials before each presentation of the 50 time-compressed speech signals included in the test proper. A 5-second interval of silence was allotted between each one of the speech signals.

Presentation and Procedure

Each *S* was tested individually on two separate occasions in a room separate from the *E*s. The first presentation consisted of 24 adaptation

trials and the 50 items from the test proper under normal speech conditions. This condition was the original tape recording (master tape). After all *Ss* received the normal run, they were assigned randomly to one of the three experimental conditions, 17 *Ss* per condition. Each *S* received only one treatment, exactly 14 days after the normal run. The normal run was given to establish a vocabulary baseline and to gather normal reaction-time data for each *S*. If *S* missed a word or words on the normal run, the data from those words were excluded from analysis on both the normal and experimental runs for that particular *S*.

The testing equipment was housed in a trailer on the school grounds. The normal and time-compressed speech signals were played back on a Roberts 770 stereo tape recorder (See Fig. 1) and received binaurally

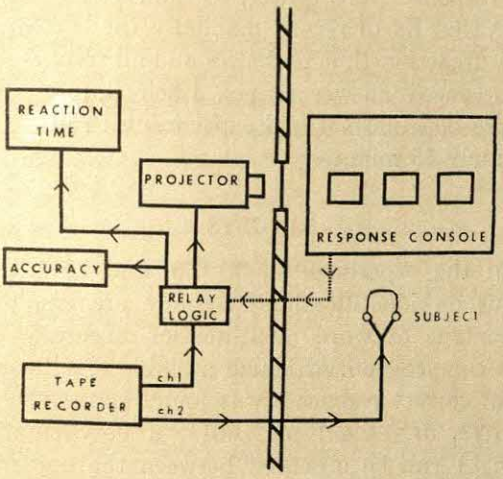


FIG. 1. Schematic block diagram of instrumentation used to measure identification accuracy and latency measures for normal and three time-compressed conditions.

by the *Ss* through TDH-39 headphones at a comfortable listening level (35–40 dB SPL re .0002 Ubar as measured with a Bruel and Kjaer-2203 Sound Level Meter and Coupler). Speech signals were transmitted from one channel of the stereo tape deck via headphones to the *Ss*. The second channel was used to trigger a voice key for visual presentation of the items and to a speaker which allowed the *Es* to aurally monitor the tape. If *S* did not respond within 3½ seconds of the allotted 5-second interval, a warning light flashed which permitted the *Es* to stop the tape and wait until a response was made.

The visual stimuli were rear projected on three panels which were in a horizontal row in front of the *Ss* at about the level of their eyes. At the

end of the phrase, "Show me—," a signal was sent via the voice key to a shutter which opened to allow the appropriate slide to be projected. After hearing the speech signal, *S* pressed the appropriate panel which best represented what he thought he heard. Correct responses were immediately reinforced automatically with an *M* and *M* candy which was delievered to a tray in front of *S*. These remained in the tray until the end of the testing session. After a response was made, either correct or incorrect, a signal was sent to the slide projector which automatically advanced to the next slide. Reaction time was measured automatically in hundredths of a second with a Haydon stop clock (Series 42). The latency counter was started by the signal which triggered the shutter on the projector and was stopped automatically as soon as a response was made. Also, the response (left, center, or right) was signaled by a pilot light which permitted *E*s to record manually the correct as well as incorrect responses. Reaction-time measures and all responses were recorded on previously prepared answer sheets. Each *S* received standardized instructions before both normal and experimental runs. A single run per *S* was approximately 15 minutes.

RESULTS

The results of the experimental, 30, 50, and 70% time-compressed speech conditions on identification accuracy are displayed in Fig. 2. The mean percentage of word identification accuracy decreased as a function of time compression. Although a highly significant decrement in the percentage of correct responses was found across the treatment conditions ($F = 77.22$, $df = 2,49$; $p < .001$), a very small decrement in intelligibility (5%) can be observed between the normal, 30 and 50% conditions (Fig. 2). A very large decrement in the percentage of correct responses occurred between the 50 and 70% conditions.

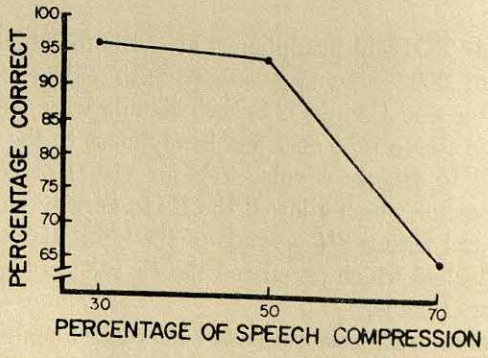


FIG. 2. Percentage correct scores of identification accuracy as a function of three conditions of time-compressed speech conditions (30, 50, and 70%).

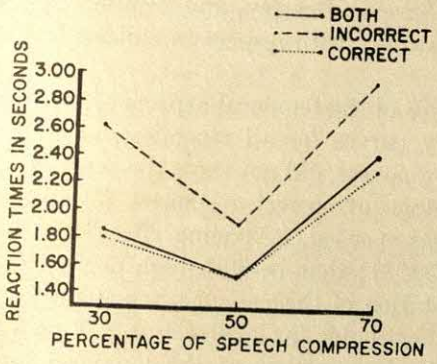


FIG. 3. Reaction times in hundredths of seconds as a function of three conditions of time-compressed speech signals (30, 50, and 70%) for *both* (incorrect and correct combined) as well as for correct and incorrect separately.

Mean latencies for *both* the incorrect and the correct responses, as well as separately for the correct and incorrect responses, are shown in Fig. 3. A significant change for *both* (incorrect and correct combined) was noted ($F = 16.04$; $df = 2,48$; $p < .001$). A Newman-Kuels test on the mean of the treatments indicated that the mean latency for the 50% treatment was significantly less than the 30 and 70% treatments, and that the mean of the 30% treatment was significantly less than the 70% condition. The same significant effect was found for the incorrect responses ($F = 5.55$; $df = 2,44$; $p < .01$); that is, the 50% treatment was significantly less than the 30 and 70% conditions, with the mean of the 30% treatment significantly less than the mean of the 70% treatment. The incorrect mean latency scores also were highly significantly different from the correct mean latency scores ($t = 266.9$; $df = 46$).[3]

DISCUSSION

The results demonstrated a significant decrement in the percentage of correct responses as a function of the treatment conditions. In Fig. 2 it can be observed, however, that very little decrement occurred in the percentage of correct responses between the normal, 30 and 50% conditions. Other investigators who used compressed speech materials with adult *Ss* established that about 30% of the temporal redundancy in speech can be removed without any appreciable decline in comprehension or speech intelligibility (Fairbanks, Guttman, and Miron, 1957; Foulke and Sticht, 1966; Zemlin, Daniloff, and Shriner, 1968; Daniloff, Shriner,

[3] The test for correlated means was used because data were obtained over all the *Ss* by taking the mean correct latency and mean error latency for a single *S*. Four *Ss* had no errors (one in 30% condition and three in 50%); consequently, their data were not included in this analysis.

and Zemlin, 1968; Shriner, Beasley, and Zemlin, 1969). The same results are reported in this study with respect to children's percentage of correct responses.

When beyond 30% of the temporal aspects of the speech signal are discarded, the latency curves for all responses, as well as for the correct and the incorrect responses, did not show the same decline in performance as did the percentage of correct responses for identification accuracy. As the message was speeded, a "pacing effect" can be observed for the 50% condition (Fig. 3); that is, with high degrees of intelligibility and with faster presentation of the message, the strategy of the *Ss* is changing. This pacing effect holds for incorrect as well as for correct responses.

Aaronson (1967), in a review of the temporal factors in perception and short-term memory, and Aaronson and Markowitz (1967) have provided evidence that stimulus factors and the strategies which are developed by *Ss* affect the temporal course of perceptual processing at an initial sensing or detection stage and at a subsequent identification or encoding stage. Broadbent (1958) has theorized a sensory storage system (S) and a perceptual system (P) to explain information processing. Broadbent's scheme can be used to interpret the pacing effect observed in this study. Measures of latency of response were taken from the offset of the stimulus. Since a direct representation of the acoustical stimulus has been postulated in the S system prior to transfer to the P system, the speech signal will be shorter for the 50% condition than for the 30% condition; that is, since the signal in the S system is shorter when compared to the 30% condition, the amount of information put into the P system is also shorter which would provide for faster retrieval and shorter latencies for the 50% condition.

With repeated presentation of the same information in less than normal time, the efficiency (strategy) of the child's signal processing system is working at maximum speed for correct responses which in turn triggers and paces his motor responses. The same processes are apparently working for the incorrect responses as well but somewhat to a lesser degree because of the uncertainty of the decision factor.

The pacing effect probably is not observed for the 70% condition because of the marked decrease in the intelligibility of the acoustic stimulus. A marked decrease in intelligibility increases reaction time measures. This particular condition supports Hecker, Stevens, and Williams' (1966) finding that reaction-time measures increase with corresponding decreases in intelligibility scores. This study attempted to directly test their contention that the inclusion of reaction-time measures may help to resolve small differences in intelligibility when combined with conventional intelligibility testing measures (assuming some degree

of independence between the two as they suggest). This could not be readily resolved on the basis of the results of this study because of the significant pacing effect. A very small decrement in identification accuracy, though, can be observed between the 30 and 50% conditions (Fig. 2) without a corresponding increase in latency, as one would probably be led to predict from the Hecker, Stevens, and Williams study. The results of a second study (Shriner, Beasley, and Zemlin, 1969) which used three conditions of frequency-divided speech signals, with and without time distortion, found that when small differences (3–4%) in identification accuracy scores occurred, then the inclusion of accompanying reaction-time measures did not appear to resolve these differences. When intelligibility begins to decline markedly, then reaction time begins to play a more sensitive role. The results of that study suggested that latency and percentage of correct responses are highly dependent.

An attempt was made to look for trends with respect to consistent consonant confusions for the words that were never missed on any of the conditions and for the words most frequently missed; however, no specific trends were noted. The reason is more a result of the PICSI test, for one would theoretically expect systematic consonant confusions based on manner and place of production. A more systematic index needs to be developed to test children's consonant confusions.

This study demonstrated that 50% of the time element of a speech signal could be removed without any appreciable effects in the degradation of intelligibility. The possibility of using time-compressed speech signals with computer-based programmed instruction becomes a distinct possibility. It also appears that an extension of this study would give valuable insight into speech signal processing behavior of normal children as well as valuable insight into children who somehow fail to meet the criteria of normalcy.

REFERENCES

AARONSON, D. Temporal factors in perception and short-term memory. *Psychological Bulletin,* 1967, **67**, 130–144.

AARONSON, D., AND MARKOWITZ, N. Immediate recall of normal and "compressed" auditory sequences. Paper presented at Eastern Psychological Association, April, 1967, Boston, Massachusetts.

BROADBENT, D. E. *Perception and Communication.* New York: Pergamon Press, 1958.

DANILOFF, R. G., SHRINER, T. H., AND ZEMLIN, W. R. The intelligibility of vowels altered in duration and frequency. *Journal of the Acoustical Society of Amercia,* 1968, **44**, 700–707.

FAIRBANKS, G., EVERITT, W. L., AND JAEGER, R. P. Methods for time and frequency compression-expansion of speech. *Transactions of the Institute of Radio Engineers, Professional Group on Audio,* 1954, February, 7–12.

Fairbanks, G., Guttman, N., and Miron, M. Effects of time compression upon the comprehension of connected speech. *Journal of Speech and Hearing Disorders,* 1957, **22,** 10–19.

Foulke, E., and Sticht, T. G. Listening rate preferences of college students for literary material of moderate difficulty. *Journal of Auditory Research,* 1966, **6,** 397–401.

Hecker, M. H. L., Stevens, K. N., and Williams, C. E. Measurements of reaction time in intelligibility tests. *Journal of the Acoustical Society of America,* 1966, **39,** 1188–1189.

Seidel, S. J. The PICSI test: Picture identification for children—a standardized index. Unpublished master's thesis, Indiana Univ., 1963.

Shriner, T. H., Beasley, D. S., and Zemlin, W. R. Effects of frequency divided speech signals on children's identification accuracy and reaction-time measures. *Journal of Speech and Hearing Research,* 1969, **12,** in press.

Zemlin, W. R., Daniloff, R. G., and Shriner, T. H. The difficulty of listening to time compressed speech. *Journal of Speech and Hearing Research,* 1969, **11,** 875–881.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 541–552 (1969)

Oddity Learning and Learning Sets in Children[1]

ANNE SARAVO
The Fels Research Institute

AND

EUGENE S. GOLLIN
University of Colorado

Three to six-year-old children were given learning set training on 1 or 11 oddity/discrimination problems, followed by a within-*S* comparison of oddity performance on random reversal and minimally interferent oddity paradigms. Most 3 year-olds were unable to maintain a stable discrimination performance. Cue reversals in the oddity test trials disrupted performance, indicating a possible confounding factor in previous studies of oddity learning. Possession of the relevant verbal labels was not related to oddity performance.

A majority of preschool-age children are unable to solve the oddity problem (Gollin and Shirk, 1966; Gollin, Saravo and Salten, 1967; Hill, 1965; Lipsitt and Serunian, 1963). The failure of 4–5 year-old *S*s in oddity learning tasks is somewhat surprising since a problem requiring a response to oddity ("point to the one that is not the same") appears on the Stanford-Binet at age 4½; and also monkeys (Meyer and Harlow, 1949), pigeons (Nevin and Niebold, 1966) and retardates (MA3-5) (House, 1964) have learned the oddity problem, albeit with prolonged training (up to 1000 trials).

Young children may have difficulty learning the oddity problem because the odd stimulus on each trial is generally a stimulus which was nonodd on a previous trial, and vice versa, constituting a reversal. This is particularly true in the standard two-pair (AA BB) oddity problem (ABB BAA). Since ease of reversal learning appears to be related to ontogenetic and phylogenetic status (Voronin, 1962; Gollin, 1964; Saravo,

[1]This research was carried out under Contract No. PH 43 65-1011 between the Public Health Service, National Institutes of Child Health and Human Development and the Fels Research Institute, E. S. Gollin, Principal Investigator. Appreciation is expressed to the staff of Mills Lawn Elementary School, Yellow Springs Community Nursery School, Dayschool of the Fairmont Presbyterian Church and Greater Springfield Day Care Center for making children available for this study.

1967; Saravo and Kolodny, 1969), the primary effects of age differences and task variables in previous studies may have been on the reversal component of the problem, not on the oddity response *per se*. Another related factor that might account for the young child's difficulty with the oddity problem is his tendency toward irrelevant or noncontingent response strategies, including both perservation and alternation to position and object cues (Hill, 1965; Schusterman, 1964). In the oddity problem, responding on the basis of specific position or object cues will not produce solution of the problem, but may produce intermittent reinforcement of these irrelevant response tendencies and thereby strengthen them.

The following experiment was designed to investigate these dimensions of interference in the oddity learning of young children. A learning set procedure was devised which would eliminate the interference produced by cue perseveration; correction techniques were used to reduce *S*'s tendency to perseverate to position. Transfer tests were administered so that oddity performance on the standard reversal oddity procedure could be compared with an oddity test that eliminated reversal.

The relationship between oddity performance and possession of the relevant verbal designators which might serve as mediators was also assessed. Kendler, Kendler, and Wells (1960) have postulated developmental differences in the ability to utilize mediational responses to control discriminative behavior; other developmental theorists (Werner, 1937; Wohlwill, 1962; Gollin, 1968) also emphasize a developmental transition between perceptual-motor and conceptual-verbal cognitive modes. Since Scott (1964) found that retarded children were able to respond to questions of sameness, but not difference, *S*s in this study were asked to identify the odd object in terms of both similarity and difference; e.g., both "not same" and "different" were included in the hope that the results would more accurately reflect concept, not vocabulary differences.

METHOD

Subjects. The *S*s were 163 children from nursery and elementary schools in the Dayton, Ohio area: 29 three-year-olds, 94 four-year-olds, and 46 six-year-old first grade children. All *S*s were experimentally naive.

Apparatus. The child sat before a 24 × 42-in. grey panel which was decorated with colored cartoon animals. The stimulus aperture, situated in the middle of the lower half of the panel, contained three 4½ × 5⅛-in. clear plastic panels arranged side by side. The stimuli were displayed in the compartments behind these panels; a masonite screen could be

lowered so as to cover the stimulus aperture. When the child pushed a panel the stimulus was displaced to reveal the presence or absence of a candy reward in a small well in the floor of the compartment. A total of 44 pairs of multi-dimensional junk objects were used as stimuli, with two identical stimuli and one odd stimulus comprising a set. Positive (odd) and negative (non-odd) members of the stimulus sets were counterbalanced by reversing these contingencies for half the *Ss*.

Design. Each *S* in the experimental training groups was trained on eleven 3-stimulus oddity discrimination problems (ABB, CDD, EFF, etc.) followed by a 30-trial oddity transfer test which included a 15-trial new set (NS) task and a 15-trial reversal-oddity (RS) task. Order of presentation was counterbalanced so that half the *Ss* were given NS first, half were presented with RS first. The original design called for 16 *Ss* in each of two experimental cells at each of two ages, 3 and 4 years. Sixty-two percent of the 3-year group failed the first problem; thus there remained only the two groups of 4-year-olds in the experiment proper. These two groups, Group RS-NS and Group NS-RS, represented the counterbalancing order of the two transfer tasks. The comparison of the two transfer tasks NS and RS was a second, within-S factor. This completed a 2×2 (Order $\times$ Transfer tasks) factorial design with repeated measures on one factor.

A second group of children was run to assess the effect of the training procedure on the transfer tasks and to compare performance on the two transfer tasks as a function of age. These *Ss* (controls) were given only the first problem in training. After controls reached criterion on the first training problem they were immediately given the transfer trials. There were four control groups of 16 *Ss* each with age (4 vs. 6 years), order (Group RS-NS vs. Group NS-RS), and transfer tasks (RS vs. NS) constituting a $2 \times 2 \times 2$ factorial with repeated measures on the transfer task variable.

Training task. Eleven combinations of 22 pairs of multidimensional (junk) objects were used to make up eleven 3-stimulus training sets, one for each problem. Within each problem, only one object was the positive, odd stimulus, so that only the first trial of each problem was an unconfounded measure of oddity learning; the remaining trials were oddity-discrimination trials. On the first trial of the first problem, *E* baited the positive stimulus with the candy reward in *S*'s sight. Correction was allowed for a maximum of five trials in the first set; thereafter and in the second and succeeding sets, a repeated trials procedure was used when *S* made an error. In this procedure, the stimuli remained in the same position on succeeding trials until *S* was correct. Only one stimulus object could be chosen on each trial presentation. Within each problem, the

position of the positive stimulus followed a random order, excluding the repeated presentation trials. In the first trial of each problem the new, odd stimulus never appeared in the same position as the odd stimulus in the preceding trial. Criterion was 9/10 correct in the first problem, and four successive correct responses on the remaining ten problems. If an *S* did not reach a criterion within 30 trials, the verbalization test was given and another *S* was substituted in the experiment.

Transfer trials. Immediately after completing the training task *S*s were given the transfer trials. The RS task (15 trials) conformed to the standard two-pair random reversal oddity problem, in which reversal or stimulus change trials are intermixed with non-change trials in a Gellerman order, e.g., ABB, BAB, BAA, BBA, BAB, etc., in order to eliminate stimulus alternation as a confounding solution strategy. The NS task (15 trials) used ten new 3-stimulus objects sets presented in a sequence directly comparable to the RS task, e.g., ABB, BAB, DCC, EEF, EFE, etc. Thus the sequences of the two tasks were exactly the same with one exception: each time a reversal occurred in the RS task, the entire stimulus set was changed in the NS task, eliminating the reversal component as a source of interference. The odd stimulus was in each of the three positions, right, middle, left, equally on each task. The noncorrection procedure was used throughout the transfer tasks.

Verbalization test. At the end of the session, *E* sat beside *S* and asked: "How did you know where the candy was? Where was it always?" After recording *S*'s answer, *E* placed two, new, different stimuli on a table in front of *S*. Holding a stimulus identical to one of those on the table, she told *S* to point to the one which was either (1) the same as, or (2) different from, or (3) not the same as the stimulus she held. These three questions were asked for each of three sets of stimuli in counterbalanced order. As *E* put away each stimulus set, she said "very good." During the verbalization trials, *E* gave no indication as to whether or not the child's verbalization was correct.

RESULTS

Training. Thirty children were dropped from the Experimental groups for failure to reach criterion on the first oddity-discrimination problem: 12 of 44 4-year-olds, and 18 of 29 3-year-old *S*s. The relative number of children unable to reach criterion was greater in the 3-year than in the 4-year group ($\chi^2 = 7.36$, $p < .01$). The 3-year-old *S*s who were terminated manifested cue perseveration tendencies. Eighty-eight percent repeated their previous error at least twice during the repeated trials correction procedure. Twenty-five percent repeated their error at least four times, while three *S*s made the same response ten or more times

consecutively. Only one 4-year-old S who was terminated showed as many as four perseverative responses.

The 4-year-old Ss improved over successive training problems. An analysis of variance on the number of errors to criterion over problems indicated this improvement was significant ($F = 2.60$, $df = 10/300$, $p < .01$).

First trial correct responses, the unconfounded measure of oddity responding, also differed over problems 2–11 (Cochran Q test, $Q = 36.79$, $df = 9$, $p < .001$) but the efficiency of oddity responding did not appear to improve in a linear fashion. With performance on the second problem as a referent, there was a significant improvement in oddity responding only on problems 5, 6, and 9 (McNemar tests for significance of change, $df = 1$; $\chi^2 = 6.05$, $p < .02$; $\chi^2 = 11.25$, $p < .001$ and $\chi^2 = 5.5$, $p < .02$, respectively). In addition, t tests between problems 2–6 and 7–11 revealed that while the number of oddity responses *decreased* from the first to the second block ($t = 2.09$, $df = 31$, $p < .05$) the number of second trial correct responses which followed an error on the first trial, the "once-wrong-then-right" response, increased. This increase in "once-wrong-then-right" responses suggested Ss had hit on a strategy which required minimum effort. In sum, while discrimination performance improved over problems, oddity responding showed some improvement midway in learning, then declined to its initial level.

Eighteen 4-year and 14 6-year control group children failed to reach criterion on the first task. There were no significant differences between the control groups or between the controls and experimentals on this measure nor on the number of errors to criterion.

Transfer tasks. Table 1 presents the means and standard deviations of the transfer errors made by nonterminated Ss. A $3 \times 2 \times 2$ mixed design analysis of variance was performed on these scores with Order (Group RS-NS vs. Group NS-RS) and Groups (Exp. 4-year vs. Control 4-year vs. Control 6-year Ss) as between S factors, and transfer Conditions (RS vs. NS) as a within-S comparison. The analysis showed that the RS task was a more difficult test of oddity than the NS task ($F = 19.02$, $df = 1/90$, $p < .001$), and that the three groups performed differently ($F = 15.17$, $df = 2/90$, $p < .001$). Separate two-group analyses indicated that the 6-year oddity performance was superior to Control 4-year Ss ($F = 24.79$, $df = 1/60$, $p < .001$) and Exp. 4-year Ss ($F = 10.84$, $df = 1/60$, $p < .01$). In addition, the performance of the Exp. 4-year group was superior to the Control 4-year group ($F = 5.63$, $df = 1/60$, $p < .05$). The reversal effect was significant in each of these subsidiary analyses ($p < .001$), with no other main effects or interactions reaching significance. These effects were congruent with a count of the

TABLE 1
NUMBER OF ERRORS IN TRANSFER TASK

Order	New set		Reversal	
	M	SD	M	SD
		Experimental 4-year-olds		
NS-RS	4.38	2.33	6.88	3.24
RS-NS	5.62	2.73	7.12	3.03
		Control 4-year-olds		
NS-RS	6.88	3.86	7.50	3.79
RS-NS	6.81	2.97	8.75	1.88
		Control 6-year-olds		
NS-RS	2.88	3.56	3.06	4.81
RS-NS	3.06	3.49	4.75	3.71

number of *S*s who reached a learning criterion of 11/15 successive correct in the transfer trials: 25% of the 4-year controls, 63% of the 6-year controls, and 56% of the 4-year experimentals were able to reach a criterion of 11/15 correct in one or both tasks. Thus, both training and age were found to relate to oddity performance in the transfer trials.

The younger children's performance was consistent with the age trend: Of the 3-year-old *S*s who were not terminated ($N = 11$), only four gave better than a chance performance in the transfer task; all of these *S*s demonstrated oddity learning during the NS task and only in the NS-RS order.

Verbalization. On the verbal designation task each *S* could receive a maximum of nine points, one for each correct identification. Figure 1 presents the mean correct responses of all *S*s on this verbalization task, including *S*s who were terminated for failure to reach criterion on the training task. An unequal N analysis of variance was performed on the verbal task scores with age (3 vs. 4 years), training performance (terminated vs. non-terminated *S*s) and verbal designator (*same, not same,* and *different*) as the variables in a 3-factor mixed design with repeated measures on the type of verbal designation. Four-year-old *S*s were able to make more correct identifications ($F = 10.89$, $df = 1/64$, $p < .01$) than the 3-year group, and the three verbal designations were of differential difficulty ($F = 22.57$, $df = 2/128$, $p < .001$). Duncan multiple range tests revealed that correctly identifying the *different* object was the most difficult, *not same* was intermediate in difficulty, while *same* was the easiest to identify (*different* vs. *same*, $p < .001$; *different* vs. *not same*, $p < .05$; *same* vs. *not same*, $p < .001$). No other effects were significant.

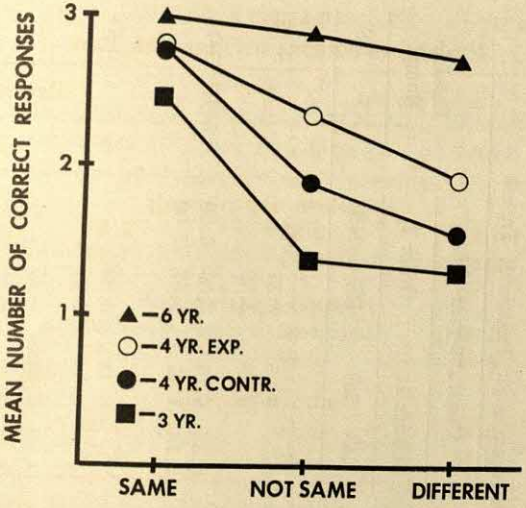


FIG. 1. Mean correct object identifications in the verbalization trials for all *Ss*.

The 6-year group was not included in this analysis because their scores would have produced heterogeneity of variance. There were no differences in the 6-year identification performance (t tests) as a function of either verbal designator or training task performance.

A number of correlations were also run on the transfer task and verbal task data. Within each age group, there were no significant correlations between performance on either or both transfer tasks and the ability to identify the objects on the basis of *same, not same,* or *different.* Thus, while possession of the oddity labels was related to age, it did not predict performance on either oddity or oddity-discrimination.

Fifty-three percent of the 6-year-old *Ss* were able to verbalize solution of the oddity problem, while only two 4-year-old *Ss* and one 3-year-old *S* could verbalize the solution. The performance of the 6-year-old *Ss* who could verbalize solution was superior on both transfer tasks to the children who were unable to verbalize solution ($F = 41.39$, $df = 1/30$, $p < .001$). However, since oddity solution was achieved by 4-year-old and nonverbalizing 6-year-old *Ss*, it can not be concluded that this verbal ability was a prerequisite for, or mediated solution of, the problem.

Nonsolution responses. Two nonsolution response tendencies, perseveration and win-stay, lose-shift (WSLS), are common in young children (Hill, 1965). For the RS task, these two response strategies may be applied to either object or position cues; in the NS task, they could only be consistently applied to position. Table 2 presents the means and

TABLE 2

MEAN NUMBER OF RESPONSES CORRESPONDING TO NONSOLUTION STRATEGIES IN TRANSFER TRIALS

	Experimental 4-year-olds				Control 4-year-olds				Control 6-year-olds				Possible Range	
	Order 1		Order 2		Order 1		Order 2		Order 1		Order 2			
	NS	RS	RS	NS	NS	RS	RS	NS	NS	RS	RS	NS	NS	RS
Perseveration														
Object M		10.56	11.19			9.69	12.00			8.25	9.31			8–15
SD		2.78	2.74			1.92	2.58			0.58	2.52			
Position M	7.06	7.56	8.12	7.56	8.12	7.81	9.12	8.19	6.38	6.19	7.06	6.69	5–15	5–15
SD	1.95	1.75	2.03	1.59	2.36	1.87	2.33	2.64	1.78	1.97	1.57	2.02		
WSLS														
Object M		7.62	7.44			7.56	7.94			6.44	6.88			0–14
SD		1.63	1.55			1.71	1.39			2.19	2.28			
Position M	3.31	6.25	6.12	5.00	5.94	6.31	7.69	5.25	2.69	2.31	4.25	2.50	0–14	0–14
SD	1.78	2.91	3.03	2.78	3.30	3.07	2.15	2.82	2.84	3.68	3.59	3.22		

standard deviations of the number of nonsolution responses made by *Ss* in each group. The range of possible scores is also included in the table for each condition. Separate analyses of variance were performed on each of these strategies with Groups, Order and Transfer tasks as the factors for the position strategies, and Groups and Order as the factors in the object strategy analyses. Table 3 presents a summary of these

TABLE 3
SUMMARY OF ANALYSES ON STRATEGY RESPONSES

	Position Perseveration		Object Perseveration		Position WSLS		Object WSLS	
	F	*df*	*F*	*df*	*F*	*df*	*F*	*df*
Groups (G)	9.49[c]	2/90	8.55[c]	2/90	14.24[c]	2/90	3.24[a]	2/90
Order (O)	—	—	7.92[b]	1/90	—	—	—	—
G × O	—	—	—	—	—	—	—	—
Transfer tasks (T)	—	—			19.81[c]	1/90		
T × G	—	—			—	—		
T × O	—	—			—	—		
T × G × O	—	—			4.44[a]	2/90		

Duncan multiple range tests

	Position Perseveration	Object Perseveration	Position WSLS	Object WSLS
Control 6-year-olds vs. Control 4-year-olds	$p \times < .001$	$p \times < .01$	$p \times < .001$	$p \times < .05$
Control 6-year-olds vs. Experimental 4-year-olds	$p \times < .001$	$p \times < .001$	$p \times < .001$	n.s.
Control 4-year-olds vs. Experimental 4-year-olds	$p \times < .05$	n.s.	$p \times < .05$	n.s.

[a] <.05.
[b] <.01.
[c] <.001.

analyses, and the Duncan multiple range tests performed on the group factors. The analysis on the number of responses corresponding to position perseveration indicated that only the effect of Groups was significant. The multiple range tests indicated that both training and age were related to the amount of positioning in the transfer trials. The 6-year group made the fewest position responses, followed by the experimental 4-year group. The control 4-year *Ss* made the most position responses.

A similar analysis of object perseveration on the RS task revealed

that more cue perseveration occurred when the reversal paradigm preceded the NS task. Age was the only reliable factor within the groups since both 4-year groups made more responses to one stimulus than did the 6-year-old *Ss*, but there were no reliable differences between the 4-year groups.

On the WSLS analysis of position responding, the differences between all groups was significant, indicating an influence of both age and training. There was also more overall WSLS positioning in the RS task than the NS task. The 6-year-old *Ss* were affected by the order of the two tasks; they made more WSLS responses in the RS task only when it preceded the NS task (Task $\times$ Order, $F = 5.86$, $df = 1/30$, $p < .05$; NS vs. RS at Order RS-NS, $t = 2.72$, $df = 15$, $p < .02$). For both 4-year-groups, WSLS positioning was affected only by type of task; there were more WSLS responses to position in the reversal task ($F = 12.50$, $df = 1/30$, $p < .01$, and $F = 5.87$, $df = 1/30$, $p < .05$, for Experimental and Control 4-year-old *Ss*, respectively).

SUMMARY OF FINDINGS AND DISCUSSION

Four-year-old children are able to learn oddity with reasonable efficiency under conditions which minimize interference. Although Harlow (1958) suggested that oddity learning was beyond the capacity of the 4-year-old child, in the present study those children who were able to form a stable discrimination indicated mastery of the oddity problem when training and testing corresponded to the NS paradigm.

Three-year-old children had greater difficulty than the other age groups in establishing a stable discrimination in the 3-stimulus oddity discrimination problem. This is consistent with previous findings (Saravo and Kolodny, 1969) in which 3-year-old *Ss* were unable to sustain criterion performance in successive discrimination problems.

Oddity discrimination training appeared to improve the ability to make a stable discrimination response more than it improved oddity responding. Since a stable discrimination response is a prerequisite to the solution of any problem which follows a consistent rule, this probably accounts for part of the training effect which was found in the transfer tasks, i.e., the superior transfer performance of the experimental 4-year-group compared to the control 4-year-group, who were not given the learning set training.

The two-pair RS procedure was more difficult than the NS procedure for *Ss*, indicating that the confounding effects of cue perseveration may have obscured the results of some earlier studies of oddity. While this study used a random oddity reversal task, the successive reversal oddity task is equally confounded. In the latter procedure, *Ss* are trained on

one set (ABB) until a stable discrimination occurs, then are trained on the reversal (BAA). This procedure continues with the same stimuli until some overall learning criterion is reached. In this method oddity learning cannot be inferred, since *S*s may learn to reverse to the cue of change in the stimulus set (e.g., Hoyt, 1960, 1962). Furthermore, many studies using multiple "sets" have defined a set as two pairs of stimuli, using the random two-pair reversal technique within each of these "sets."

In addition, the breakdown of oddity responding in the reversal task (i.e., in the NS-RS order) indicates that negative instrumental transfer is an important factor even when the relevant observing responses have been learned.

While overall performance as measured by errors showed no order effects, object perseveration and WSLS positioning were apparently so affected. These nonsolution responses indicated that performance was retarded when the RS task came at the beginning of oddity learning. The RS task lends itself easily to object perseveration; in the NS task object perseveration is prevented, and solution responses can be reinforced. The object perseveration tendency would therefore be reduced when the NS task preceded the RS task.

An explanation for the order effect in the WSLS position strategy for the 6-year-old *S*s is not so readily apparent. *S*s are able to respond to position in both tasks so the above explanation is not adequate. A tentative explanation is that the rapid changes in stimulus sets during the NS task increased the novelty or attention value of the object dimensions and so reduced position tendencies in the subsequent RS task. If this analysis is correct, it suggests that the position response is much more labile in 6-year-old *S*s than in younger *S*s, confirming the findings of a previous study (Gollin, Saravo, and Salten, 1967).

Oddity discrimination training produced a reduction of position responding which apparently permitted *S*s to respond to the object dimension in terms of oddity. The performance of the 6-year-old *S*s, however, indicated that such procedures were unnecessary for the older child: Nonsolution responses to both object and position were very low, even lower than the 4-year-olds who were given the training task. Thus while younger children are capable of responding to oddity, their performance is largely dependent on the extent to which the task minimizes irrelevant response tendencies. Older *S*s are not so dependent on the task structure and are able to quickly relinquish response patterns which do not produce task solution.

The correct use of the verbal designators: *same, not same,* and *different,* were of increasing difficulty for the 3- and 4-year-old *S*s but the 6-year-old *S*s performed equally well on the three designators. While the

ability to verbalize the specific solution of the oddity problem was related to superior oddity performance in the 6-year-old *Ss*, possession of the relevant concepts, *same, not same,* and *different,* was not related to oddity learning. Furthermore, the performance of the 4-year-old *Ss* indicated that neither possession of relevant verbal labels nor the verbalization of solution were necessary conditions for oddity solution.

REFERENCES

Gollin, E. S. Reversal learning and conditional discrimination in children. *Journal of Comparative and Physiological Psychology,* 1964, **58,** 441–445.

Gollin, E. S. Conditions that facilitate or impede cognitive functioning. *In* Hess, R. D. and Bear, R. M. (Eds.) *Early Education.* Chicago: Aldine Publishing Co., 1968.

Gollin, E. S., and Shirk, E. J. A developmental study of oddity-problem learning in young children. *Child Development,* 1966, **37,** 213–217.

Gollin, E. S., Saravo, A., and Salten, C. Perceptual distinctiveness and oddity-problem solving in children. *Journal of Experimental Child Psychology,* 1967, **5,** 586–596.

Harlow, H. The evolution of learning. *In* Roe, A. and Simpson, G. G. (Eds.) *Behavior and Evolution.* New Haven: Yale University Press, 1958.

Hill, S. D. The performance of young children on three discrimination—learning tasks. *Child Development,* 1965, **36,** 425–435.

Hoyt, J. M. Effect of similarity of reversal cues on learning of successive stimulus reversals in children. Unpublished Master's thesis, State University of Iowa, 1960.

Hoyt, J. M. Serial reversal and conditional discrimination learning in children. Unpublished doctoral dissertation, State University of Iowa, 1962.

House, B. J. Oddity performance in retardates: I. Acquisition and transfer. *Child Development,* 1964, **35,** 635–643.

Kendler, T. S., Kendler, H. H., and Wells, D. Reversal and non-reversal shifts in nursery school children. *Journal of Comparative Physiological Psychology,* 1960, **53,** 83–88.

Lipsitt, L. P., and Serunian, S. A. Oddity problem learning in young children. *Child Development,* 1963, **34,** 201–206.

Meyer, D. R., and Harlow, H. F. The development of transfer of response to patterning by monkeys. *Journal of Comparative Physiological Psychology,* 1949, **42,** 454–462.

Nevin, J. A., and Niebold, K. Stimulus control of matching and oddity in a pigeon. *Psychonomic Science,* 1966, **5,** 351–352.

Saravo, A., and Kolodny, M. Learning set and shift behavior in children. *Journal of Experimental Child Psychology,* 1969, **7,** 21–30.

Scott, K. G. A comparison of similarity and oddity. *Journal of Experimental Child Psychology,* 1964, **1,** 123–124.

Zeaman, D., and House, B. J. An attention theory of retardate discrimination learning. *In* Ellis, N. R. (Ed.) *Handbook of Mental Deficiency.* McGraw-Hill, 1963, 159–223.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 553–565 (1969)

The Developmental Control of Operant Motor Responding by Verbal Operants

DONALD MEICHENBAUM AND JOSEPH GOODMAN
University of Waterloo, Ontario, Canada

The developmental control of operant motor responding (tapping response) by the verbal operants "faster" and "slower" was examined with 54 kindergarten and 30 first grade children. The effects of the verbal operants on motor behavior were assessed under three conditions: *E* verbalized the verbal operants; *S* said the verbal operants to themselves aloud; *S* whispered the verbal operants using only lip movements (approximating a covert condition). The major finding was an interaction between age and mode of delivery in controlling motor behavior. Kindergarten children's motor performance approximated first grade children when self-verbalizations were overt, but covert self-instructions had minimal functional control over motor behavior. First-grade children's self-verbalizations had more functional control over motor behavior when covert than overt. The results suggested a developmental sequence of the functional interaction between speech-for-self and nonverbal operants.

Children can often be observed talking to themselves. This apparently spontaneous speech-for-self, usually of no communicative significance, often has significant behavioral consequences. Most actions do not appear to be impulsively performed and are often foreshadowed by internal verbalizations and thoughts (Vygotsky, 1962). The development of the functional interaction between language and nonverbal behavior is an intriguing topic.

Two general strategies have been employed to assess the influence of self-verbalizations on behavior. The first strategy is characterized by the *S*'s performance on a task and *E*'s subsequent inference as to the presence or absence of mediational processes. Reese (1962) has reviewed the developmental research arising from mediation theory (e.g., reversal–nonreversal learning, transposition, discrimination set), which indicates that there is a deficiency in mediation in young children compared with older children. In general, the results indicate that with age the tendency to make covert verbal orienting and representational responses increases and that these responses serve to mediate the execution of the task. It is suggested that the overt level responses in younger children have the same effect on learning as do covert responses in older children.

The other strategy, which is designed to assess the interaction between verbal and nonverbal behavior, directly manipulates the child's verbalizations and examines resulting changes in motor behavior. The child's verbalizations have been manipulated by *E* giving the child general instructions or a set, or by manipulating the content of the child's self-instructions. The ability to regulate nonverbal behavior by means of self-instructions has been examined by Luria (1961), Lovaas (1964), Birch (1967) and Bem (1967). Vygotsky (1962) and Luria (1959, 1961) have proposed a developmental sequence by which speech-for-self becomes internalized and directive of one's behavior. According to Luria, the child under two years of age is not able to use speech to direct his own behavior. However, the speech of others can initiate, direct, and control the child's behavior, but cannot inhibit or stop an on going behavior. During a second stage, the child can use his own speech to control his behavior to some extent. Luria suggests that the child's speech has a motor component which helps initiate motor behavior, but will not inhibit it regardless of the semantic content of the speech. It is only in the third stage that the semantic content of the child's speech-for-self becomes dominant, directive, and internalized. Birch (1967), using a hand lever depression task to the stimulus presentations of two lights which signified "push" and "don't push," has offered data in support of Luria's hypotheses. Lovaas (1964), working within an operant framework, found that verbal operants or the discriminative stimuli for these operants have a directing influence on motor behavior. Children were trained to verbally discriminate with the word "fast" to one stimulus light and the word "slow" to a different colored light and subsequently their rate of lever pressing and latency of response under alternate presentations of these lights was examined. Children aged five to six did not bring their operant motor responding under stimulus control. In pretraining the *S*'s rate of responding had to be brought under the control of *E*'s verbal commands before the *S*'s own verbal behavior could exercise control over his nonverbal behavior. Bem (1967) assessed verbal self-control with three year olds by requiring them to press a lever that number of times corresponding to the number of lights displayed which was subsequently covered. She trained the three year olds to respond correctly in the presence of external feedback and then shifted control of their behavior to their own self-instructions by means of a fading procedure. The training procedure included verbal training, verbal fading, motor training, verbal-motor training, and verbal-motor fading.

Lovaas' and Bem's results suggest a developmental sequence by which overt verbalizations of an adult or *E*, followed by the child's overt self-verbalizations, followed by covert self-verbalizations result in verbal

control of motor behavior. The present study was designed to examine the relative efficacy of the differential modes of delivery of verbalizations in governing nonverbal behavior. In order to assess the degree of verbal control of motor behavior, a finger tapping task similar to Lovaas (1964) was used in which the child was instructed to tap while the words "faster" and "slower" were verbalized by *E*; or verbalized by the *S* aloud; or whispered by *S* using only lip movements. The latter condition approximates that of a covert condition.[1] The amount of verbal control of motor behavior is evidenced by the degree of change in the rate of tapping.

In summary, the relative effectiveness of the mode of delivery of verbalizations (covert vs. overt vs. externally administered) in controlling tapping behavior was examined for two age levels (kindergarten and first grade). The influence of the content of the verbal operant "faster" and "slower" on motor behavior was examined as a within subject variable.

METHOD

Subjects

The *S*s were 54 kindergarten (27 females, 27 males) and 30 first-grade children (14 females, 16 males) who were in a public elementary school. The *S*s were randomly assigned, equating for sex, to each of three experimental conditions.

Procedure and Apparatus

Each child was taken by the *E* to a testing room in his school. The room was equipped with a child's table and chairs. Prior to the administration of the tapping task *E* assessed the child's understanding of the words "faster" and "slower." The child was required to clap his hands faster and slower and then identify which of two toy trucks and which of two clock-like dials placed on a vertical board was going faster and slower. A criterion of four correct consecutive trials on each of these tasks was required in order for *S* to continue. All *S*s were able to meet this criterion quite readily. Following the preassessment, which indicated that the *S* understood the meaning of the words faster and slower, the finger tapping task was presented.

[1] For the purposes of the present paper the whisper condition has been labeled a covert condition since many of the children only moved their lips and did not vocalize the words. One may rather conceptualize the aloud and whisper conditions as two different types of overt responding, one with lip movements and sounds and the other with lip movements only.

The finger tapper consisted of a metal telegraph key with a black plastic knob (one inch in diameter). The key was mounted on a $6 \times 5 \times 1$-inch black wooden base and connected to a counter. The *E* pressed the key several times and invited *S* to try it. *S* was shown the proper method of tapping, namely, with the middle and index fingers while resting the other fingers on the wooden base.

The *S*'s were divided equally among the following experimental conditions: *Externally administered* condition in which the child finger tapped while the verbal operants were spoken by *E; Overt* condition in which the child finger tapped while the verbal operants were said aloud by the child; *Covert* condition in which the child finger tapped while the verbal operants were said quietly by the child (lip movements). Under each of the experimental conditions the *S* was run through six phases: (1) an operant phase; (2) a phase where the effect of the word "letter" on tapping speed was assessed; followed by the verbal operants (3) "faster" and (4) "slower"; (5) "letter" and finally (6) another operant phase. Every child was required to press the finger tapper for two trials of 15 seconds with a 15-second rest between trials. Thus, all *S*s had 12 tapping trials, namely, two operants, two "letter," two "faster," two "slower," two "letter," and two operants. The only difference between experimental groups was that the words "letter, faster, and slower" were presented externally, overtly, or covertly.

The operant period provided a baseline measure of the child's tapping speed when there was no explicit task demand as to the rate of response. The first "letter" condition was presented next in order to accustom the *S*'s to the concomitant production of two operants, tapping and self-verbalization, and to assess the interference effect of verbalizing on finger tapping speed. The letter condition was repeated after "faster and slower" in order to further assess the interference effect of verbalizations on motor behavior. The effect of the words "faster" and "slower" on tapping speed reflect the degree of control of such verbal operants on a motor operant. The order of the words "faster" and "slower" was counterbalanced. The final operant period was included to assess any fatigue effect on tapping behavior.

Before each trial *S*s in the overt condition were given the following instructions: "I want you to tap, and all the while that you are tapping you are to say the word (letter, faster, slower) over and over again. Say it out loud." Before each phase the *E* modeled the appropriate procedure for *S* and had the *S* briefly practice it. Thus, before each trial the *S* was instructed to say the word out loud while tapping. *S*s in the covert condition were given the same set of instructions, but were told to "say the word so softly that I can't even hear you, like this." *E* modeled

lip movements for the appropriate word. Before each trial the *S* was instructed to whisper the word while tapping. The rate of self-verbalizations for the overt and covert conditions was unspecified, and each *S* emitted the verbal operants at his own rate. The rate of verbalizations by *E* in the externally administered condition was yoked to the *S*'s rate in the other two conditions. The mean rate *E* emitted the verbal operants was 25 times per 15-second interval. The *S*s under the "faster" and "slower" phases for all three experimental conditions were instructed to tap "the way the word means" and was reminded that he had successfully identified the faster and slower toy trucks and clock dials.

In summary, all *S*s were required to tap twice under each of an operant, letter, faster, slower, letter, and operant phase. The independent variables were the mode of delivery of verbalizations (i.e., whether the words "letter, faster, and slower" were verbalized by *E*, by *S* aloud, or by *S* quietly by means of lip movements) and age of the *S* (kindergarten vs. first grade). The major dependent variable is the total tapping speed for the two trials under each of the above phases.

RESULTS

The results are examined separately for each of the phases of the experiment, namely, the operant phases, the "letter" phases, and the phase of the verbal operants of "faster" and "slower." The analyses presented are for the mean total tapping performance for both trials, since analyses for the separate trials revealed the same general set of findings. No significant differences in tapping speed were found for sex or for the order of presentation of the words "faster" and "slower" and these factors are not considered further in the analyses.

Operant phases. Two operant phases were included in the experiment. The first operant phase provided a baseline measurement of tapping speeds for the two grades of children. A significant difference ($p < .05$) was found between kindergarten and first grade children in their mean total operant speeds, the respective means were 72.17 ($SD = 4.85$) and 88.07 ($SD = 5.20$). The increased speed of the first grade children was also found on the final operant phase which was included to assess a fatigue effect. The kindergarten children had a mean total tapping speed of 75.79 ($SD = 8.50$) and the first graders a mean of 89.70 ($SD = 7.65$). The observed age differences in tapping speed are consistent with evidence on growth trends in motor coordination and motor skills (Thompson, 1962). No evidence for a fatigue effect was found suggesting that any changes in rate of tapping speed observed are a function of the verbal operants and not of within-subject changes such as fatigue. No significant differences were found between the three experimental groups

during both operant phases indicating that the groups were equated for initial tapping speeds and when the differential treatments were removed the groups tapping speeds were once again comparable. Figs. 1 and 2 compare the three experimental conditions separately for kindergarten and first grade children. Table 1 and Fig. 3 each compares directly the kindergarten and first grade children for each of the three experimental conditions.

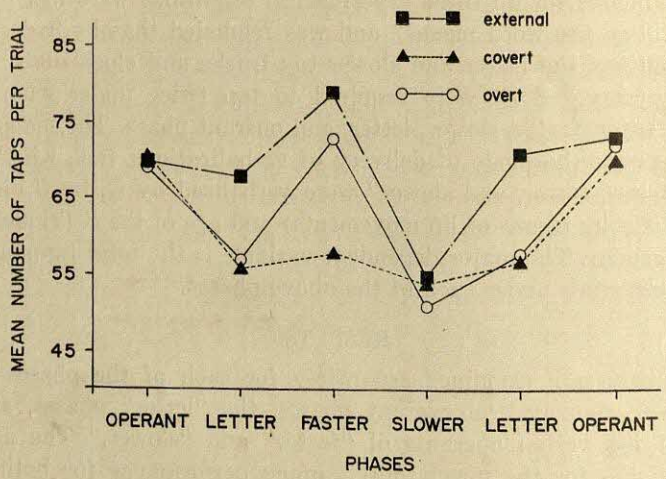


FIG. 1. Mean number of taps per trial for external, overt, and covert conditions for kindergarten children.

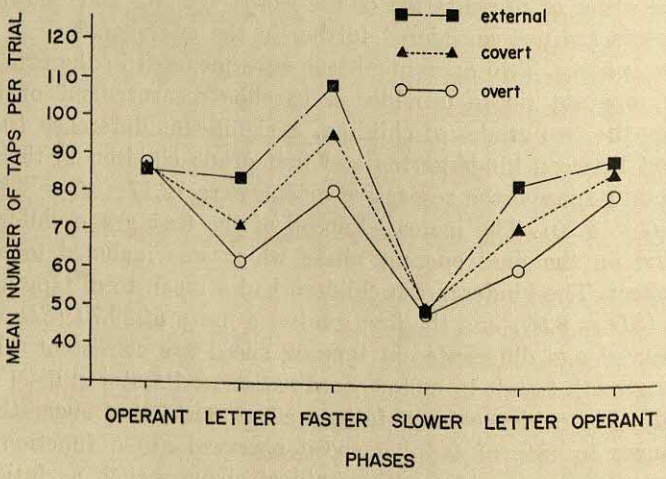


FIG. 2. Mean number of taps per trial for external, overt, and covert conditions fo first grade children.

TABLE 1

MEAN TAPPING PERFORMANCE FOR KINDERGARTEN AND FIRST GRADERS FOR EXTERNAL, OVERT, AND COVERT CONDITIONS

	Kindergarten			First Grade		
	External	Overt	Covert	External	Overt	Covert
Letter $\bar{X}$	71.47	57.50	56.15	84.20	62.60	72.50
SD	13.49	7.91	12.64	21.35	12.39	12.34
Faster $\bar{X}$	79.75	73.50	59.75	109.30	81.00	96.30
SD	15.42	18.56	17.33	18.16	24.84	28.57
Slower $\bar{X}$	55.10	50.90	54.05	50.70	49.80	51.10
SD	17.16	12.99	14.21	14.87	14.24	9.99
Letter $\bar{X}$	71.87	58.90	58.20	83.90	61.40	72.20
SD	13.80	11.88	13.03	19.93	16.77	14.88

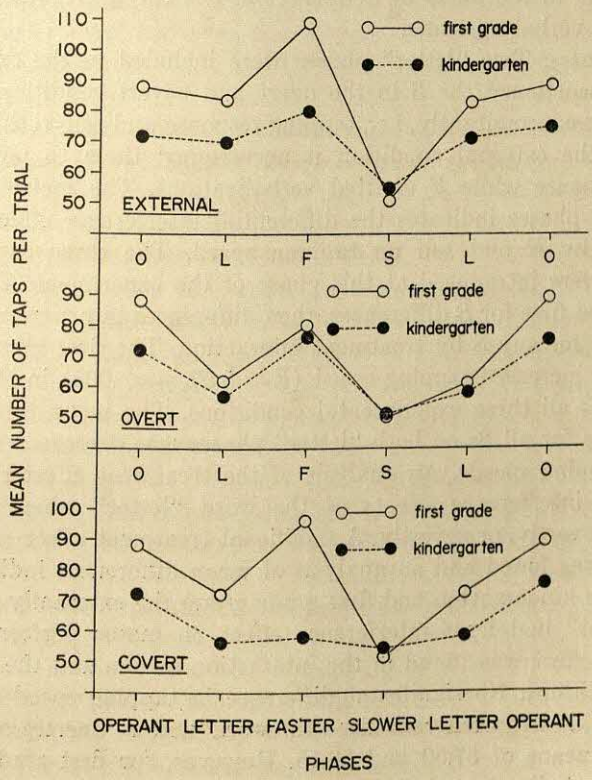


FIG. 3. Comparison of mean number of taps per trial between kindergarten and first grade children for external, overt, and covert conditions.

Before examining the effects of the mode of delivery of the verbal operants (external, overt, or covert) on motor behavior, the frequency of the verbal presentations in each condition was examined. Unfortunately, the records of the emission rates of the verbal operants for the *Ss* in this study were lost because of mechanical difficulty. However, Meichenbaum and Goodman (1968) used the same finger tapping task with a comparable population ($N = 30$). No significant difference was found between the rate of emission of the verbal operants in the overt and covert conditions. The mean emission rates for overtly produced verbal operants was 23.50 ($SD = 8.50$) and for covertly produced operants 21.75 ($SD = 9.45$). The externally administered verbalizations had been yoked to the performance of the *Ss* in the overt and covert condition. Generalizing from the *Ss* in the second study it appears that the emission rates of verbal operants is constant across overt and covert conditions and thus any observed differences in tapping performance would be due to the mode of delivery and not the differential emission rates of the verbal operants.

Letter phases. Two "letter" phases were included in the experiment. The first accustomed the *S* in the overt and covert conditions to emit two operants concomitantly, i.e., tapping response and self-verbalizations. For *Ss* in the external condition it accustomed them to perform the tapping response while *E* emitted verbalizations. The inclusion of the two "letter" phases indicates the differential interference effects of verbalizations by *E* and self on tapping speed. The three experimental conditions were introduced at this phase of the experiment. The results are examined first for *S* differences, then differences in experimental conditions, and for any *S* by treatment interaction. The first graders maintained their increased tapping speed ($F = 5.57$, $p < .005$) in the "letter" phase across all three experimental conditions. The mean total tapping performance for all *Ss* on both "letter" phases was decreased from their operant tapping speeds. An analysis of the treatment effect reveals the differential interference effects of the word "letter" when verbalized externally, overtly, or covertly. A significant treatment effect ($F = 11.59$, $p < .001$) was found and an analysis of mean differences indicated that for both the kindergarten and first grade group the externally verbalized word "letter" had least interference effect on motor performance. An interesting effect was found in the interaction of age and the overt and covert conditions. No significant difference in tapping speed was found when *Ss* in kindergarten verbalized the word "letter" overtly or covertly, respective means of 57.50 and 56.15. However, for first-grade children the overt verbalization of the word "letter" interfered significantly more with tapping speed than did the covert expression of the word. The mean

for the overt condition was 62.60 and for the covert condition 72.50. The exact same relationships were observed on the second "letter" phase providing greater reliability to the differential interference effect as a function of mode of delivery. In summary, the data from the "letter" phase indicates that (a) first-grade children tap faster than kindergarten children; (b) the externally administered verbalization has least interference effect; and (c) there is no differential interference for kindergarten children when self-verbalizing the word "letter" overtly or covertly, but for first grade children the overt expression of the word interferes more with performance.

Faster phase. A significant *S* difference ($F = 14.49$, $p < .001$) was evidenced in the "faster" phase with the kindergarten children tapping significantly slower than first-grade children when combining across the three experimental groups. However, a separate analysis of the mean total tapping speed for each of the experimental conditions indicates that differences between kindergarten and first grade children are significant for only the external and covert conditions, but not the overt condition. Kindergarten children, when overtly self-instructing themselves to tap, did not significantly differ from the first-grade children. A significant treatment effect ($F = 6.19$, $p < .004$) and interaction effect between treatment by *S*s ($F = 2.84$, $p < .05$) was found. An analysis of the mean difference revealed that for kindergarten *S*s the covert condition was least efficient in increasing tapping speed and was significantly different from the external and overt conditions which did not differ significantly from each other. For the first-grade children the most efficient condition was the external, then covert, but they did not significantly differ from each other. The overt or aloud condition for first-grade children had least control over the nonverbal operant. An analysis of difference scores where performance on the initial operant or performance for the "letter" condition are controlled revealed the same set of relationships. In summary, the data from the "faster" phase indicates that (a) first-grade children tap faster than kindergarten children, but this difference is only significant under external and covert conditions, not the overt condition; (b) the differential effectiveness of the verbal operant "faster" in modifying the nonverbal operant tapping response is a function of the mode of delivery of the verbalization and the age of the child. For first-grade children the increase in the rate of tapping was equally effected when the verbalization "faster" was externally or covertly administered, and had least functional significance when the self-verbalization was overt; whereas for the kindergarten children the external and overt conditions had equal functional control in modifying tapping speed, but the covert condition had least functional control.

Slower phase. The differential effectiveness of the verbal operant "slower" on the nonverbal operant was examined as a function of the mode of delivery and age. An analysis of the mean total tapping performance reveals the absence of a subject, treatment, or interaction effect. The Ss did not differ in the absolute level of tapping speed, but if one examines the difference scores between tapping performance in the "slower" phase and the preceding phase (either "letter" or "faster" because of the counterbalancing) then an interesting set of relationships is revealed. The use of difference scores assumes that when the *S* was instructed to tap "slower" he used his previous tapping performance as a baseline in order to decrease his speed. Table 2 indicates the mean difference in total tapping speeds between "letter" phase and "slower" phase, and between "faster" and "slower" phases. The larger the number, the greater the decrease in tapping speed under the "slower" phase from the previous phase. An analysis of both the letter minus slower, and faster minus slower phases indicates that first graders are decreasing their tapping performance significantly more ($p < .01$) than kindergarten children. However, a separate analysis of the relative speed of tapping for kindergarten vs. first-grade children for each of the three experimental conditions indicates that the difference in tapping speed between groups becomes nonsignificant for the overt condition. The same finding was observed for the "faster" phase. Kindergarten children when self-instructing themselves aloud to increase or decrease tapping speed do not significantly differ from first grade children. A significant treatment effect and interaction effect for both "letter minus slower" and "faster minus slower" was obtained (p values less than .05). An analysis of mean differences revealed that for kindergarten children the covert condition was most inefficient in modifying speed of tapping responses and the external condition was most efficient, but not significantly different from the overt condition. For first grade children the most efficient condition was the external, then covert and finally overt. All three experimental conditions differed significantly ($p < .05$) from each other for first-grade children. The aloud or overt self-verbalization had least functional control over nonverbal operant for first-grade children. In summary, the data from the "slower" phase indicates (a) no significant difference between groups or treatments when the absolute level of tapping responses are examined; (b) if the previous level of tapping performance is included by means of difference scores, then the same general pattern of findings as in the "faster" phase is evident. For kindergarten children the verbal operant "slower" has less functional control of motor behavior when the mode of delivery is covert than external or overt. When the self-verbalization is overt the performance

TABLE 2

MEAN DIFFERENCE IN TOTAL TAPPING SPEED BETWEEN "LETTER AND SLOWER" PHASES AND "FASTER AND SLOWER" PHASES[a]

	External	Overt	Covert
Letter minus Slower			
Kindergarten	16.40	8.40	2.10
First Grade	33.50	11.80	20.90
Faster minus Slower			
Kindergarten	24.30	23.20	5.70
First Grade	58.70	30.20	45.20

[a] The larger the number the greater the decrease in tapping speed under the "slower" phase from the previous phase.

of kindergarten children approximates and is nonsignificant from that of first graders. First-grade children manifest most control when verbal operants are external, then covert; whereas overt self-verbalizations result in the least functional control of motor behavior.

DISCUSSION

The results indicate that the mode of delivery of verbalizations is an important variable in the development of verbal control of motor behavior. An interaction between the age of the child and the mode of delivery revealed that (a) Kindergarten children's motor performance approximated that of first grade children when self-verbalizations were aloud or overt, but covert self-verbalizations had minimal functional control over nonverbal operants; (b) first-grade children's self-verbalizations had more functional significance when covert rather than overt. First-grade children's motor behavior was significantly more responsive to *E*'s external commands than that of kindergarten children. The above relationships were evident for both the verbal operants "faster" and "slower" and for the general interference effect of the word "letter." The differences between the experimental conditions were not due to differences in motor ability, differential understanding of the words, or fatigue effects. The groups had been equated for tapping speed and were carefully tested as to their comprehension of the verbal operants.

Meichenbaum and Goodman (1968) used the same finger tapping task to examine the relationship between Kagan's (1966) reflection-impulsivity dimension and the verbal control of motor behavior with kindergarten children. A significant relationship between the manner (number and pattern) in which the child produced verbal operants to control his motor behavior and the conceptual tempo dimension of reflection-impulsivity was found. The reflective kindergarten children were significantly more likely to use their self-verbalizations in a semantic

manner, tapping several times for each self-instruction; whereas impulsive children were more likely to use self-instructions in a motoric manner, tapping each time they produced a verbalization. These results indicate that children use self-verbalizations in different manners to control their nonverbal behaviors and this may relate to the development of cognitive styles.

The results of the present study are consistent with the general finding that for younger children covert self-verbalizations have minimal functional significance, but their overt self-instructions seem to serve the same function as older children's covert verbalizations (Reese, 1962; Kendler and Kendler, 1962; Birch, 1966). An interesting finding was that for first graders overt self-verbalizations had less functional control over motor behavior than covert self-verbalizations. Kendler, Kendler, and Carrick (1966) using a inferential problem solution task also found that an overt label facilitated performance of kindergarten children, but interfered with the performance of third grade children. These findings suggest that language gains functional significance by means of a developmental sequence: First motor behavior is brought under the control of an adult's or *E*'s overt verbalizations; then under the child's overt self-verbalization and then under a diminished self-verbalization such as talking quietly to oneself as in the present experiment, and finally under the control of implicit self-verbalizations. The results of the present study suggest that for the task employed, the kindergarten children were at the stage where overt self-verbalizations had greater functional control than covert self-verbalizations; whereas first graders passed this stage. Forcing first-grade children to make overt their self-verbalizations resulted in less control of motor behavior.

What factors contribute to developmental verbal control of motor behavior? Lovaas (1964) has suggested such factors as the deferred imitation of parents' verbalizations, the reinforcement for behaving in accordance with one's self-verbalizations, and the self-reinforcement arising from the performance of nonverbal behavior which is in accord with previous self-verbalizations (e.g., planning behavior). Present research is being conducted to examine the role of such social-developmental factors in explaining the functional interaction between verbal and nonverbal behavior.

One of the pedagogical implications of the present study is that teaching procedures employed be adaptive to the child's optimal stage of self-verbalization. The present results suggest that younger children be encouraged to self-verbalize aloud and this should be gradually faded to implicit speech. Overt speech-for-self appears to be a constructive and facilitative operant which has important behavioral consequences.

REFERENCES

BEM, SANDRA L. Verbal self-control: The establishment of effective self-instruction. *Journal of Experimental Psychology,* 1967, **74,** 485–491.

BIRCH, D. Verbal control of nonverbal behavior. *Journal of Experimental Child Psychology,* 1966, **4,** 266–275.

BIRCH, D. Some effects of the verbal system on the nonverbal behavior of preschool children. Paper presented at the Society for Research in Child Development, 1967.

KAGAN, J. Impulsive and reflective children: Significance of conceptual tempo. In J. D. Krumboltz (Ed.). *Learning and the educational process.* Chicago: Rand McNally, 1965.

KENDLER, H. H., AND KENDLER, TRACY S. Vertical and horizontal processes in problem solving. *Psychological Review,* 1962, **69,** 1–16.

KENDLER, TRACY S., KENDLER, H. H., AND CARRICK, MARTHA A. Verbal labels and inferential problem solution of children. *Child Development,* 1966, **37,** 749–763.

LOVAAS, O. I. Cue properties of words: The control of operant responding by rate and content of verbal operants. *Child Development,* 1964, **35,** 245–256.

LURIA, A. R. The directive function of speech in development. *Word,* 1959, **15,** 341–352.

LURIA, A. R. *The role of speech in the regulation of normal and abnormal behavior.* New York: Liveright, 1961.

MEICHENBAUM, D., AND GOODMAN, J. Reflection-impulsivity and verbal control of motor behavior. *Child Development,* in press, 1969.

REESE, H. W. Verbal mediation as a function of age level. *Psychological Bulletin,* 1962, **59,** 502–509.

TERRACE, H. S. Errorless transfer of a discrimination across two continua. *Journal of Experimental Analysis of Behavior,* 1963, **6,** 223–232.

THOMPSON, G. *Child psychology.* Boston: Houghton-Mifflin, 1962.

VYGOTSKY, L. S. *Thought and language.* New York: Wiley, 1962.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 566–577 (1969)

Perceptual Variables in the Transposition Behavior of Children[1]

THOMAS J. TIGHE AND LOUISE S. TIGHE
Dartmouth College

Two experiments on transposition behavior in first-grade children are reported. In the first, perceptual pretraining with either 3 stimulus values appearing on the dimension in pretraining or with continuous variation of the dimension was found to significantly facilitate transposition. But identical perceptual pretraining with only 2 stimulus values on the dimension had no effect on transposition, regardless of whether these values were relatively close or distant on the continuum. The second experiment compared 3 modes of stimulus presentation during the transposition test phase: (1) Simultaneous presentation of training and test pairs; (2) alternating presentation of training and test pairs; (3) presentation of the test pair only, as in the customary transposition test procedure. Transposition was maximal in (1), minimal in (3), and intermediate in (2).

This paper presents two experiments in a series of studies of the role of perceptual learning in the discrimination processes of children. Earlier experiments have defined a set of pretraining conditions of a perceptual nature which produce facilitation of discrimination reversal learning (Tighe, 1965; Tighe and Tighe, 1968a) and transposition (Tighe and Tighe, 1968b) in young children. The major feature of the pretraining is that *S*s are required to make nonreinforced same-different judgments to successively presented stimulus objects which vary along the dimensions (size and brightness) appearing in the subsequent discrimination tasks. Following the conceptions of differentiation theory (Gibson and Gibson, 1955; Tighe and Tighe, 1966), which provided the impetus for these experiments, it is assumed that such pretraining promotes differentiation of the task stimuli and that facilitation of discrimination following such treatment is due to an increase in *S*'s sensitivity to the distinguishing features of the task.

An interesting finding in these experiments is that facilitation of reversal obtained when four stimulus values appeared on each dimension

[1] This research was supported by Public Health Service Research Grant MH11088, from the National Institute of Mental Health. The authors are grateful for the generous cooperation of Gordon Tate, Superintendant of the Lebanon, N. H. school system and for the assistance of the principles and teachers of this school system.

in pretraining but not when only two stimulus values specified each dimension. Appropriate controls established that this outcome was not due to the greater number of stimulus exposures or longer training time required by the four-value as compared to the two-value condition. The differential effect of two- vs. four-values per dimension was confirmed in a later study (Tighe and Tighe, 1968c) which also demonstrated that reversal learning is facilitated when only three stimulus values appear on each dimension in pretraining. Experiment I tests two hypotheses which might account for these data. The first hypothesis is that degree of perceived continuity among stimulus values may be a critical factor in the isolation of dimensions of this type. That is, when three or four stimuli, as opposed to two stimuli, specified each dimension, the children may have been more likely to perceive similarity and therefore more likely to respond to them as a distinct class of cues. The second hypothesis is that the smaller distance between stimulus values in the three- and four-value conditions evoked greater perceptual search and thus made more likely the isolation of the task dimensions. These hypotheses are tested by comparing the transposition behavior on the dimension of height of children who were perceptually pretrained under the following conditions of dimensional representation. (1) Two-value-far condition: Only two stimulus values (A and C) specified the height dimension. (2) Three-value condition: Stimuli A and C were presented plus a third intermediate value B. (3) Two-value-near condition: Two relatively close stimulus values were presented. One-half the *S*s in this condition were pretrained with stimuli A and B and the other half with stimuli B and C. (4) Continuous variation condition: Subjects in this condition experienced continuous variation of the dimension between the extremes of stimulus values A and C.

METHOD

Subjects. The *S*s were 50 first-grade children divided into 5 groups of 10 each for differential pretraining treatment. Each *S* was run individually in a small room adjacent to his classroom.

Pretraining procedure and apparatus. In all pretraining conditions, the stimuli were 1⅝-inch diameter wood cylinders which varied in height. All cylinders were covered with black enamel paint. The stimuli were presented in an apparatus designed to enable discrete presentation of stimuli and to shield the motions of *E* from *S*'s view. For pretraining conditions other than the continuous variation condition, this consisted of a four-walled plywood structure 18 (long) × 9 (wide) × 19-inches (high) which was placed on a table lengthwise before *S*. The front and rear walls of the apparatus contained sliding doors which could be opened

by *E* from behind the apparatus. At the start of each trial the front door was opened to reveal a three-walled compartment 8 inches deep with walls 8 (wide) × 19 inches (high). Cut into the floor of this compartment were two ½ (deep) × 1¾-inches diameter holes which served as stimulus holders. One of these holes, intended for placement of comparison (Co) stimuli, was in the approximate center of the compartment floor while the other, intended for placement of a standard (St) stimulus, was placed about two inches to the right and forward of this point.

In the two-value-far training condition *E* presented only a 2½-inch-high cylinder (stimulus A) or a 7-inch-high cylinder (stimulus C). On a given presentation series, one of these stimuli was designated as a St and placed in the appropriate hole. The *E* then opened the front door of the compartment and said, "Let's pretend that this object is yours. Look at it carefully. I am going to show you some others one at a time. You tell me when you see one that is exactly the same as yours." The *E* then closed the sliding door, removed the St, placed a Co stimulus, and reopened the door with the question, "Is this exactly the same as yours?" This question was repeated with each Co stimulus presented. Both stimuli were used as standards and each child received four series of stimulus presentations with each of the two standards. A series consisted of presentation of the St and the 2 possible Co stimuli, i.e., stimuli A and C, and *S* rendered a same-different judgment to each stimulus presented. The order of selection of the standards and the order of presentation of the Co stimuli were random.

In the three-value condition *E* presented stimuli A and C plus a 4¾-inch high cylinder (stimulus B). Stimuli A and C served as the standards and in all respects the training was identical to that in the two-value condition except that a series consisted of a St plus the three possible Co stimuli, i.e., stimuli A, B, and C.

In the two-value-near condition, training was identical to the two-value-far condition except that ½ the *S*s saw stimuli A and B while the other half saw only stimuli B and C.

In the continuous variation condition the stimulus presentation apparatus was constructed around a 32-inch-high table with a 9 × 28-inch top. A 28 × 35-inch piece of plywood attached at a right angle to the front of the table shielded the table and motions of *E* from *S*'s view. A sliding door in the top center of the shield could be opened to reveal a three-walled compartment 8-inches deep with walls 8 (wide) × 16 inches (high). The rear wall of the compartment was a sliding door which could be manipulated by *E* from behind the apparatus. A moveable metal rod projected up through the center of a circular opening 1¾-inch in diameter in the center of the compartment floor. The tip of

this rod was inserted into the center of the base of a 13¾-inch-high cylinder which, so placed, extended through the hole and below the floor of the compartment. By manipulating a set screw beneath the floor of the compartment *E* could raise or lower the rod, thus changing the height of the cylinder visible to *S*. About 2 inches to the right and forward from this variable cylinder a ½-inch deep and 1¾-inch-diameter hole served as a holder for the St stimuli. Stimuli A and C were used as the standards. On a given presentation series, the sliding door was opened far enough to permit *S* to view the St but not the variable stimulus, and *E* said, "Let's pretend that this object is yours. Look at it carefully. I am going to take this away and show you another one and I want you to tell me when it looks exactly the same as yours." The *E* closed the sliding door, removed the St, set the Co stimulus to either height A or C (i.e., to the height which differed from whatever St had been presented), and reopened the door. The *E* then varied the height of the Co so that it continuously approached the height of the St, and instructed *S* to say "stop" when the Co looked exactly the same as "his." A series consisted of presentation of a St and one adjustment of the variable cylinder. In this way, *S* had the same number of exposures to each *stationary* stimulus value during pretraining as *S*s in the other pretraining conditions, i.e., on each series *S* saw as stationary stimuli the St, the initial value of the Co, and the terminal value of the Co. As in the other pretraining conditions, four series were given with each St stimulus.

None of the judgments or observations in any pretraining condition were corrected or reinforced and all were based solely on vision. The number of experiences with each Co and St stimulus was the same in all pretraining conditions. The time required for perceptual pretraining was about the same in the various conditions, ranging from 5 to 7 minutes. In order to make clear to *S* the meaning of the instructions he was to receive in pretraining, *S*s in all pretraining treatments were first given about 4 minutes practice in making corrected same-different judgments on cookie cutters varying in shape and hue.

Ten additional *S*s worked for an equivalent period of time on tasks unrelated to the subsequent discrimination tasks and designed to control for nonspecific transfer effects. The tasks were modifications of the Picture Completion and Picture Arrangement subtests of the Wechsler Intelligence Scale for Children.

Discrimination tasks. Immediately following pretraining all *S*s learned a discrimination between a pair of black cylinders 4 inches and 5½-inches in height. These stimuli were presented on a simple turntable device as described in Kendler and Kendler (1959) except that the rear of the presentation device was extended in height to provide a uniform

background for the training and test stimuli. The *S* made his choice by picking up one of the cylinders on each trial and if correct found a marble under it. The *S* was told that the "game" was to see how soon he could find a marble every time he chose. Tall and short were used equally often as the positive stimulus within all treatments. When *S* made 9 correct out of ten consecutive responses, the transposition task was presented immediately without change of instructions. All *S*s received ten presentations of 10¾ versus 15-inch-high cylinders with all choices reinforced. These stimulus values were selected as being likely to yield little transposition in nonpretrained *S*s of this age and therefore as being likely to produce a suitable baseline against which to measure the effect of perceptual pretraining. Selection of these values was guided by Zeiler's data (1967). Throughout training the stimulus pairs were presented according to a prearranged sequence designed to control for order and position effects. Each *S* was assigned to groups and positive stimuli on a predetermined random basis. At the end of the experiment *S* was allowed to choose a prize from an assortment which included charms, bubble gum, whistles, raisins, jack sets, Tootsie Rolls, pencils, pencil sharpeners, M & M's, and flutes.

RESULTS

The various pretraining treatments did not differentially affect speed of learning the initial discrimination. Table 1 gives the mean number of trials to criterion in this task by each treatment group. Analysis of variance of these data yielded $F < 1$. In analyzing the data of the transposition test, an *S* was considered to have exhibited transposition if he choose on a relational basis on at least 8 out of the 10 transposition test trials, an outcome which would be expected to occur by chance approximately 5 times in 100. Table 1 shows the number of *S*s in each

TABLE 1

MEAN NUMBER OF TRIALS TO CRITERION IN INITIAL DISCRIMINATION, MEAN NUMBER OF TRANSPOSITION RESPONSES, AND NUMBER OF TRANSPOSERS IN EACH PRETRAINING TREATMENT

Treatment	Mean number of trials to criterion	Mean number of transposition responses	Number of transposers
Control	7.5	4.0	1
Two-value-far	8.9	5.6	1
Two-value-near	5.2	5.2	3
Three-value	8.6	9.3	9
Continuous	7.4	10.0	10

treatment group falling in this category. Comparisons of groups on this measure were made by Fisher Exact Tests. The procedure outlined by Ryan (1959) was used in assigning a level of significance to these multiple comparisons. Since there are a total of ten possible comparisons, a p level of .005 is required for each comparison to reduce the error rate per experiment to the .05 level. By this criterion, the continuous-variation condition produced significantly more transposers than the control condition ($p < .001$) as did the three-value condition ($p < .005$). Comparison of the control condition with the two-value-near condition did not approach significance ($p = .29$). In other comparisons of interest, more transposition obtained in the continuous-variation condition than in the two-value-near condition ($p < .005$), and the difference between the latter condition and the three-value condition approached significance ($p = .01$). The identical pattern of conclusions resulted from application of Wilcoxon's test for unpaired replicates to the differences between treatment groups in frequency of transposition (relational) responses. The mean number of transposition responses for each group is listed in Table 1.

Thus, perceptual pretraining with three stimulus values per dimension or with continuous variation between two values significantly facilitated transposition while pretraining with two values per dimension, whether relatively near or distant on the continuum, had no effect on transposition behavior.

EXPERIMENT II

The differentiation theory of perceptual development posits that sensitivity to stimulation is not a static affair but changes significantly with *S*'s perceptual experience. Of particular importance for the study of discrimination is the assumption that the ability to abstract distinguishing features (dimensions) of stimulation is often a matter of learning, and that we can expect to find both developmental and individual differences in the level of such perceptual learning. An analysis of transposition behavior from this viewpoint suggests that the reported increase in transposition with age (Kuenne, 1946; Alberts and Ehrenfreund, 1951) reflects an increase in *S*'s tendency to identify the stimulus objects of the experiment as members of a common continuum. In the typical unidimensional transposition experiment the relation between the stimulus elements of the training pair or between the elements of the test pair is probably relatively easy to perceive since the two stimulus values in each of these pairs are generally quite close on the continuum. But it may be difficult for the young child to classify the training pair and the test pair as belonging to the same continuum since these pairs usually

lie far apart. If training and test pairs are responded to as essentially different classes of stimuli then it is unlikely that the response strategy followed in training will generalize to the test phase. In support of this view is the observation that marked developmental differences in transposition behavior usually obtain only when the test stimuli are remote from the training stimuli (e.g., Kuenne, 1946).

If failure to transpose reflects failure to perceive the training and test stimuli as members of a common continuum, then we would expect that transposition in the young child would be facilitated by a procedure which allowed the child to simultaneously view the training and test stimuli *during the test for transposition,* as opposed to the customary test procedure in which training and test pairs are presented in succession. As a test of this hypothesis, first-grade children received discrimination training on a pair of stimuli differing in height and then underwent remote transposition tests under one of the following conditions: (1) Simultaneous presentation of the training and test pairs during the test phase; (2) alternating presentation of the training and test pairs during the test phase; (3) presentation of the test pair only, i.e., the customary transposition test procedure.

This analysis should be contrasted with that of Stevenson and Bitterman (1955), which is also concerned with the manner in which young *S*s perceive the task stimuli. They distinguish two relational processes in transposition behavior—one that is abstract and one that is tied to the absolute properties of the specific training situation (e.g., the region of the afferent continuum). The former process is assumed to characterize the learning of the mature *S,* while the latter process is thought to prevail in the younger child and in infrahuman *S*s. Under this view, transposition in young children is seen as resulting from failure to discriminate between the training and test sets. In contrast to the differentiation view, then, the Stevenson–Bitterman hypothesis predicts that little transposition will obtain under the simultaneous presentation test condition of the present experiment since this condition should assure discrimination between the training and test sets. However, Stevenson and Bitterman might argue that the simultaneous presentation condition is sufficient to instate the "abstract" process assumed to mediate transposition in mature *S*s.

Method

Subjects. The *S*s were 36 first-grade children divided into three groups of 12 each. Each child was tested individually in a room adjacent to his classroom.

Apparatus. An apparatus similar to that employed in the two-value

and three-value pretraining conditions of Experiment I was used in both the discrimination and transposition test phases. The apparatus was a four-walled plywood structure 36 (long) × 9 (wide) × 19-inches (high), and it was placed on a table lengthwise before *S*. The front and rear walls of the apparatus each contained two sliding doors which opened from the center and which were operated by *E* from behind the apparatus. When both of the front doors were opened they revealed a three-walled compartment 8 (deep) × 16 inches (wide) with walls 19-inches high. Cut into the floor of this compartment were four ½ (deep) × 1¾-inch-diameter holes which served as stimulus holders. These holes were aligned about 3 inches from center to center with their front edges about 2 inches from the front of the compartment. A 4 (high) × 6-inches (deep) gray cardboard partition ran from the front edge to the rear of the apparatus between the two center holes. Thus, when only one of the sliding doors was opened *S* saw a three-walled compartment 8 (deep) × 8 inches (wide) with the partition forming one "wall" of the compartment.

Training procedure. During discrimination training only the last-described presentation compartment was used, i.e., at the start of each trial only one of the front sliding doors was opened. All groups were trained to discriminate between 4 versus 5½-inch-high black cylinders. The procedure and instructions were those used in the discrimination task of Experiment I. Tall and short were used equally often as the positive stimulus within all training treatments. When *S* made 9 correct out of 10 consecutive responses, the transposition tests were presented immediately. For all groups the transposition test stimuli consisted of 10¾ versus 15-inch-high black cylinders. Subjects who received alternating presentation of the training pair and the test pair during transposition test (Alternating Group) and *S*s who received presentation of the test pair only (Successive Group) viewed the test stimuli in the same presentation compartment utilized in training. For *S*s who received simultaneous presentation of training and test pairs (Simultaneous Group), both sliding doors on the front of the apparatus were opened to reveal the four stimuli aligned in the placement holes. Subjects in the Successive Group received 10 test trials each of which consisted of presentation of the test pair only. Subjects in the Alternating Group received 20 trials during the test phase consisting of 10 presentations of the test pair and 10 interspersed presentations of the training pair. Temporal order of presentation of training and test pairs was random with the restrictions that neither pair was presented more than twice in succession and that the test pair was presented on trial one. Subjects in the Simultaneous Group received 10 trials consisting of simultaneous

presentation of training and test stimuli. For these *Ss* left vs. right positioning of training and test pairs from trial to trial was random with the restriction that neither pair appeared on the same side of the central partition more than twice in succession. Subjects in the Alternating and Simultaneous Groups were reinforced for choice of either of the training stimuli and *Ss* in all groups were reinforced for choosing either member of the test pair on all test trials. At the introduction of the test stimuli in the Alternating and Successive conditions, *E* said, "Pick one of these." During Simultaneous presentation *E* designated either the left or right hand pair of stimuli and said, "Choose one of these." When *S* had made his choice, *E* pointed to the other pair and said, "Now pick one of these." The order in which the training and test pair were designated on any presentation was randomly determined with the restriction that the test pair was the first to be designated on the initial presentation of the testing series. No other instructions or information were given *S* during the testing phase of any condition. Each *S* was assigned to groups and positive stimuli on a predetermined random basis. At the end of the experiment *S* was allowed to choose a prize from the assortment described under Experiment I.

Results

There were no differences among the treatment groups in speed of learning the initial discrimination ($F < 1$). As in Experiment I, an *S* was considered to have exhibited transposition if he chose on a relational basis on 8 or more of the 10 test trials. All of the 12 *Ss* in the Simultaneous Group exhibited transposition by this measure, while 6 *Ss* transposed in the Alternating Group and 1 *S* transposed in the Successive Group. Comparisons of groups on this measure were made by the Fisher Exact Test with a p level of .016 required for each comparison to reduce the error rate per experiment to the .05 level (Ryan, 1959). By this criterion, the Simultaneous Group differed from both the Successive Group ($p < .001$) and the Alternating Group ($p < .01$). The difference between the Successive and Alternating Groups approached significance ($p = .034$). The mean number of transposition responses in the Simultaneous, Alternating, and Successive conditions were 9.9, 7.5, and 5.6, respectively. Application of Wilcoxon's test for unpaired replicates to the differences in frequency of transposition responses between groups produced the same pattern of conclusions described above.

DISCUSSION

Experiment I shows that the relations between conditions of dimensional representation in pretraining and transposition behavior are the

same as those previously observed with reversal learning in *Ss* of the same age. Perceptual pretraining with three or more stimulus values per dimension facilitates later transposition and reversal while pretraining with only two values per dimension has no effect on performance in either task. This pattern of findings suggests that a fruitful line of investigation would be the study of known parameters of transposition in relation to reversal learning and vice-versa.

The outcome of Experiment I fails to support the hypothesis that distance between stimulus values is the critical factor in the differential effect of number of values per dimension, and the fact that the facilitating effect of three values per dimension was equal to that of continuous variation indicates that the perception of continuousness *per se* is not the crucial aspect. The results of Experiment II support the preexperimental hypothesis that transposition is likely when *S* classifies training and test stimuli together in that maximum transposition occurred under simultaneous presentation of training and test pairs, least transposition occurred under successive presentation, and alternating presentation produced an intermediate amount of transposition. If identification of training and test stimuli as belonging to the same continuum is the critical factor in transposition, the question is raised as to how the effective pretraining operations of Experiment I altered *S*'s sensitivity in this manner. One possibility is suggested by the observation of Riley, McKee, and Hadley (1964) that considerable transposition obtains in young children when the stimuli involved are ones which these *Ss* find relatively easy to place in order along their dimensional attribute (e.g., different auditory intensities), but little transposition obtains when the stimuli are relatively difficult to order (e.g., different auditory frequencies). In relating this observation to Experiment I, note that it is only when three or more stimulus values appear on each dimension that stimulus ordering by *S* is likely to be brought into play. That is, only when at least three values appear would correct judgment be appreciably facilitated by location of the position of each object presented within an ordered series. This analysis suggests that pretraining with continuous dimensional variation was also effective because it served to identify the stimulus objects as members of an ordered series. Presumably, such stimulus ordering experience increases *S*'s sensitivity to the continuum relating the training and test stimuli. Supporting this analysis is the subsidiary observation of Riley, McKee, and Hadley (1964) that their visual and auditory ordering pretests probably increased frequency of transposition—an effect they attributed to a "certain instructional quality of the ordering procedures."

The present analysis of transposition behavior should be considered

in relation to the considerable number of studies comparing the effect of simultaneous vs. successive presentation of the training stimuli in relation to subsequent transposition tests in both human and infrahuman subjects (see Hebert and Krantz (1965) for a review of these experiments). The general stimulus to these studies has been the absolute-relational controversy. The argument is that if organisms learn to discriminate on the basis of the relations between the stimuli then simultaneous presentation should yield better transposition than successive presentation, while if the effective stimuli are the absolute properties of the positive and negative stimuli then presumably mode of presentation should have little effect on transposition scores. In contrast, the present argument is that the crucial question in relation to the transposition experiment is not whether organisms relate the members *within* the stimulus pairs, but whether they relate the training and test pairs. Pursuing this line of reasoning, it is questionable whether *S*'s choice behavior during transposition tests is an adequate basis upon which to infer response to absolute vs. relational properties of the stimuli during the training phase. Thus, *S*s might discriminate both the training stimuli and the test stimuli on the basis of their relative properties, yet choose different aspects of the relation (i.e., the larger vs. the smaller) during training and testing if the distinguishing feature common to these phases has not been fully abstracted.

REFERENCES

Alberts, E., and Ehrenfreund, D. Transposition in children as a function of age. *Journal of Experimental Psychology,* 1951, **41**, 30–38.

Gibson, J. J., and Gibson, E. J. Perceptual learning: Differentiation or enrichment? *Psychological Review,* 1955, **62**, 32–41.

Hebert, J. A., and Krantz, D. L. Transposition: A re-evaluation. *Psychological Bulletin,* 1965, **63**, 244–257.

Kendler, T. S., and Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *Journal of Experimental Psychology,* 1959, **58**, 56–60.

Kuenne, M. R. Experimental investigation of the relation of language to transposition behavior in young children. *Journal of Experimental Psychology,* 1946, **36**, 471–490.

Riley, D. A., McKee, J. P., and Hadley, R. W. Prediction of auditory discrimination learning and transposition from children's auditory ordering ability. *Journal of Experimental Psychology,* 1964, **67**, 324–329.

Ryan, T. A. Multiple comparisons in psychological research. *Psychological Bulletin,* 1959, **56**, 26–47.

Stevenson, H. W., and Bitterman, M. E. The distance-effect in the transposition of intermediate size by children. *American Journal of Psychology,* 1955, **68**, 274–279.

Tighe, L. S. Effect of perceptual pretraining on reversal and nonreversal shifts. *Journal of Experimental Psychology,* 1965, **70**, 379–385.

TIGHE, L. S., AND TIGHE, T. J. Discrimination learning: Two views in historical perspective. *Psychological Bulletin,* 1966, **66,** 353–370.

TIGHE, L. S., AND TIGHE, T. J. Transfer from perceptual pretraining as a function of number of stimulus values per dimension. *Psychonomic Science,* 1968c, **12,** 135–136.

TIGHE, T. J., AND TIGHE, L. S. Perceptual learning in the discrimination processes of children: An analysis of five variables in perceptual pretraining. *Journal of Experimental Psychology,* 1968a, **77,** 125–134.

TIGHE, T. J., AND TIGHE, L. S. Facilitation of transposition and reversal learning of children by prior common perceptual training. Paper presented at meetings of the Eastern Psychological Association, Washington, D. C., April, 1968b.

ZEILER, M. D. Stimulus definition and choice. *In* L. P. Lipsett and C. Spiker (Eds.). *Advances in child development and behavior.* Vol. 3. New York: Academic Press, 1967. Pp. 125–156.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 7, 578–584 (1969)

Effects of Varying Delay of Reinforcement and Postreinforcement Intervals on Learning of Retardates[1]

STUART M. KEELEY
Bowling Green State University

AND

ROBERT L. SPRAGUE
University of Illinois

All combinations of two levels of delay of reinforcement (0 and 15 seconds) and two levels of postreinforcement intervals (1 and 16 seconds) were used in a two choice discrimination learning task with institutionalized retardates. Significantly more correct responses were given under 0-second delay. Latency of responding was significantly shorter under 0-second delay than under 15-second delay, and was also shorter under 1-second postreinforcement than under 16-second postreinforcement conditions. The interaction of postreinforcement interval by blocks of trials produced a significantly greater shortening of latency in the 1-second condition over the five blocks per day. The latency results indicated that delay interval should not be confounded with postreinforcement interval since both intervals influence latency of responding.

Several studies have compared discrimination learning performance in children under delay of reinforcement conditions with performance under immediate reinforcement. With fourth grade children as *Ss* and colored lights as stimuli, investigators have found no effect of delay on acquisition performance (except in one difficult task) but have found significantly shorter response latencies to immediate reinforcement trials (Erickson and Lipsitt, 1960; Etzel and Wright, 1964; Hockman and Lipsitt, 1961). These studies suggest that the main effect of delay of reinforcement is on speed of response, not accuracy.

Slower acquisition and shorter response latencies during delay of reinforcement are found when kindergarten and first grade children serve as *Ss* and the stimuli vary in size and form (Terrell and Ware, 1961; Terrell and Ware, 1963; Ware and Terrell, 1961). Slower acqui-

[1]This study was supported in part by PHS grant NB 07346 from the National Institute of Mental Health.

sition under delay was also found by Markowitz and Renner (1966) with third graders as *S*s and line drawings as stimuli.

Discrimination learning in the mentally retarded child has been found to be deficient to that of normal children (House and Zeaman, 1958; Girardeau, 1959). Studies by Shoelkopf and Orlando (1965), Hetherington, Ross, and Pick (1964), and Ross, Hetherington, and Wray (1965) have found that reinforcement delay is an important factor in the determination of discrimination learning efficiency in retardates. Delays as brief as 5 seconds significantly reduce learning performance. These studies have not measured speed of response.

The above results indicate that the finding of a temporal gradient varies with the subject population and the complexity of the stimuli. In addition, these findings are difficult to interpret theoretically, since they have typically ignored intertrial interval as a possible important influence on discrimination performance (Terrell, 1965). Brackbill (1964) found that an interpolated task *following* immediate reinforcement impaired acquisition, suggesting that the children needed time to digest their "knowledge of results." With adults, Bourne and Bunderson (1963) found a significant postreinforcement effect in a concept formation task. As postreinforcement interval increased, the number of correct responses increased linearly. In a thesis, Wooley (1968) obtained a significant effect on latencies when she varied both delay and postfeedback in a delayed retention design similar to that used by Brackbill and Kappy (1962).

These findings suggest that an *S*'s behavior following reinforcement and/or feedback may be an important determinant of discrimination learning performance. Typically, intertrial interval is confounded with postreinforcement interval. When intertrial interval is held constant, postreinforcement interval must vary with changes in the delay of reinforcement interval. Separation of pre- and postreinforcement influences appears to be a necessary step toward an adequate theoretical explanation of the temporal gradient. The present investigation systematically manipulates pre- and postreinforcement intervals in a multiple simultaneous discrimination task with mentally retarded children. In addition to accuracy measures, latency measures, which might be expected to reflect delay influences not reflected directly by accuracy measures alone, are obtained.

METHOD

Subjects

Twelve mentally retarded boys with a mean age of 142.9 months ($SD = 18.0$) and mean mental age of 71.6 months ($SD = 9.3$) from

Lincoln State School, Lincoln, Illinois, were selected to participate in this study.[2]

Apparatus

An automatic version of the Wisconsin General Test Apparatus was used in this study. The stimuli were colored slides of two pictures taken from children's books, and the slides were projected onto the rear of two Lenscreen windows which were mounted in a panel in one wall of the experimental room. *S* was seated facing this panel, and he pressed the windows directly to indicate his response. Throughout the study whenever *S* pressed a window, the stimuli were turned off until the onset of the next stimulus presentation. Midway between the windows and slightly below them was a clear plastic container into which M & M's were ejected. The candy was easily visible to the *S* and readily obtainable throughout the experimental session. The timing sequences and the order of the stimuli were automatically programmed with standard behavioral relay equipment. Correct responses and latency of the response in tenths of a second were manually recorded by *E*s.

Procedure

Extensive pretraining was given in an attempt to minimize learning set effects and to insure that the *S*s understood their task. Pretraining was conducted over 2 days with three problems. The first problem consisted of a picture and a blank. On this problem and all subsequent problems, the correct stimuli appeared in either the left or right window as dictated by a Gellerman series. This problem was shown for 25 trials. The second problem contained a picture and an amorphous mass; this problem was presented for 25 trials. In both the first and second pretraining problems, the picture was the positive stimulus. The third and final pretraining problem was similar to the two choice discrimination problems used in the experimental sessions. An entry criterion of 20 correct responses in 25 consecutive trials was established on this third pretraining problem. All *S*s met this criterion. After meeting the criterion, 40 overlearning trials were given on the third problem.

Upon completion of pretraining, *S* began the 4-day sequence of experimental sessions. Within each session, 5 problems were presented for 10 trials each for a total of 50 trials per session. After 10 trials of a problem, *S* was given 45 seconds of rest before starting the next problem.

[2] The authors thank Dr. Louis Belinson, Superintendent, and Mr. William R. Chambers, Psychologist, for their permission to use the facilities and patients at Lincoln State School.

Two levels of delay of reinforcement (0 and 15 seconds), and two levels of postreinforcement intervals (1 and 16 seconds) were used which determined the length of a trial. The postreinforcement intervals were set 1 second longer than the delay intervals to permit the projector to change slides. Only one combination of the four possible combinations of delay and postreinforcement were given on a particular experimental day; e.g., 0-second delay and 16-second postreinforcement. On subsequent days, *S* received other combinations until he received all four possible combinations. Twelve possible orders over days of the four combinations were randomly assigned to *S*s so that no two *S*s had the same order of treatments over the four days.

On each trial *E*s recorded *S*'s responses, but latencies in tenths of a second were recorded only for the total of a 10-trial problem.

RESULTS

An analysis of variance was performed on the logarithmic transformation of the number of correct responses per problem. (A logarithmic transformation was done to reduce the heterogeneity.) Only the delay variable is significant; $F(1,11) = 5.42$, $p < .05$. The arithmetic mean number of correct responses in the 10 trial problem in the 0-second delay was 9.0, and in the 15-second delay condition was 7.6.

An analysis of variance was also performed on the reciprocal transformation of the latency data. The mean latencies in seconds for the delay and postreinforcement conditions and their sub-groups is given in Table 1. It can be readily ascertained that a significant delay effect [$F(1,11) = 43.16$, $p < .001$] is due to longer latencies under the 15-second delay condition and that the significant postreinforcement effect [$F(1,11) = 14.46$, $p < .01$] is attributable to the longer latencies under the 16-second postreinforcement condition. A significant interaction was not obtained between the delay and postreinforcement variables. Also, the means of the 15- and 1- and 0- and 16-second combinations do not

TABLE 1
MEAN LATENCIES IN SECONDS

	Delay		
	0	15	Average
Postreinforcement 1	16.2	22.0	19.1
Postreinforcement 16	18.7	25.8	22.2
Average	17.4	23.9	

TABLE 2
MEAN LATENCIES FOR A PROBLEM IN SECONDS

	Problems				
	1	2	3	4	5
Postreinforcement 1	21.0	18.9	19.2	18.7	18.0
Postreinforcement 16	20.9	23.2	22.1	23.2	21.9
Average	20.9	21.0	20.6	20.8	20.0

differ significantly. A significant difference between these means would indicate that the delay interval has a decremental effect not attributable entirely to the intertrial interval.

There is a significant problem by postreinforcement interval interaction on the latencies; $F(4,44) = 3.18$, $p < .05$. Table 2 presents the mean latencies for the two postreinforcement intervals over the five problems. It can be seen that the latencies are almost the same on the first problem, but decrease over problems in the 1-second postreinforcement condition while showing increasing trends over problems for the 16-second condition.

DISCUSSION

The significant effect of delayed reward in decreasing the number of correct responses made by mentally retarded *Ss* in a simultaneous discrimination problem agrees with the results of several other studies (Ross *et al.*, 1965; Shoelkopf and Orlando, 1965). The results indicate that retarded children have difficulty bridging the 15-second gap between response and reinforcement.

Post reinforcement interval had no effect on the number of correct responses, which may be reflecting the very simple nature of the task. However, both pre- and postreinforcement intervals did significantly affect response latency. Indeed, the briefer the interval between trials, the shorter the response latency. This finding is the primary contribution of the study. Two aspects of the latency results are important: (a) The latency measure seems to be a more sensitive index of the effects of pre- and postreinforcement intervals in that the latency measure was statistically significant on both the delay and postreinforcement factors. In addition, the accuracy data on the number of correct responses is quite prone to either floor or ceiling effects dependent upon the level of difficulty for the population being used. Latency data, by its nature being a continuous variable rather than a discrete variable (correct vs. incorrect), is less prone to such floor and ceiling effects. (b) If latency

data is collected in future delay of reinforcement studies, our results would indicate that the delay intervals should not be confounded with the postreinforcement intervals because both intervals independently influence latency of responding. This consideration seems especially relevant to competing response explanations of delay of reinforcement effects on latency and may, in part, account for the discrepancy in latency results between such studies as Hockman and Lipsitt (1961) and Etzel and Wright (1964), the former supporting a competing response explanation of delay effects, the latter not supporting it.

The nature of our results suggests that in this type of experiment with these *S*s, speed of response may be a function of the tempo of the trials. The tempo, in essence, may pace the *S*. The faster the trials are presented, the quicker are the responses and vice versa. Indication that tempo may be an important variable in determining speed of responding comes from a study by Shriner and Sprague (1969). In this investigation, a triad of stimuli were visually presented in three windows, and *S* pressed one of the windows in response to an auditory command, "Show me ————," presented by a tape recorder over headphones. When the instructions were presented under time compressed speech commands, which of course sounds to the listener as if the speaker is talking very rapidly, there was a significant decrease in latency of responding with increased time compression. Although the experimental operations differ markedly in the two studies, and altering of the stimulus characteristics cannot be equated with varying intertrial interval, both studies are suggestive of effects on latency from the tempo of the task.

The finding of significant pre- and postreinforcement effects on latencies in this investigation indicates that intertrial interval is an important consideration in delayed reward research. This study is limited to only two values of pre- and postreinforcement intervals. More definitive relationships would, of course, require the use of many intervals.

REFERENCES

Bourne, L. E., Jr., and Bunderson, C. V. Effects of delay of information feedback and length of post-feedback interval on concept identification. *Journal of Experimental Psychology*, 1963, **65**, 1–5.

Brackbill, Y. The impairment of learning under immediate reinforcement. *Journal of Experimental Child Psychology*, 1964, **1**, 199–207.

Brackbill, Y., and Kappy, M. S. Delay of reinforcement and retention. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 14–18.

Erickson, M. T., and Lipsitt, L. P. Effects of delayed reward on simultaneous and successive discrimination learning in children. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 256–260.

Etzel, B. C., and Wright, E. S. Effects of delayed reinforcement on response

latency and acquisition learning under simultaneous and successive discrimination learning in children. *Journal of Experimental Child Psychology,* 1964, **1,** 281–293.

GIRARDEAU, F. L. The formation of discrimination learning sets in mongoloid and normal children. *Journal of Comparative and Psysiological Psychology,* 1959, **52,** 566–570.

HETHERINGTON, E. M., ROSS, L. E., AND PICK, H. L., JR. Delay of reward and learning in mentally retarded and normal children. *Child Development* 1964, **35,** 653–659.

HOCKMAN, C. H., AND LIPSITT, L. P. Delay of reward gradients in discrimination learning with children for two levels of difficulty. *Journal of Comparative and Physiological Psychology,* 1961, **54,** 24–27.

HOUSE, B. J., AND ZEAMAN, D. A comparison of discrimination learning in normal and mentally defective children. *Child Development,* 1958, **29,** 411–416.

MARKOWITZ, N., AND RENNER, K. E. Feedback and the delay-retention effect. *Journal of Experimental Psychology,* 1966, **72,** 452–455.

ROSS, L. E., HETHERINGTON, M., AND WRAY, N. P. Delay of reward and the learning of a size problem by normal and retarded children. *Child Development,* 1965, **36,** 509–517.

SCHOELKOPF, A. M., AND ORLANDO, R. Delayed vs. immediate reinforcement in simultaneous discrimination problems with mentally retarded children. *Psychological Record,* 1965, **15,** 15–23.

SHRINER, T. R., AND SPRAGUE, R. L. Effects of time-compressed speech signals on children's identification accuracy and latency measures. *Journal of Experimental Child Psychology,* 1969, **7,** 532–540.

TERRELL, G. Delayed reinforcement effects. *In* L. P. Lipsitt and C. C. Spiker (Eds.), *Advances in Child Development and Behavior,* Vol. 2, pp. 127–158. New York: Academic Press, 1965.

TERRELL, G., AND WARE, R. Role of delay of reward in speed of size and form discrimination learning in childhood. *Child Development,* 1961, **32,** 409–415.

TERRELL, G., AND WARE, R. Emotionality as a function of delay of reward. *Child Development,* 1963, **34,** 495–501.

WARE, R., AND TERRELL, G. Effects of delayed reinforcement on associative and incentive factors. *Child Development,* 1961, **32,** 789–793.

WOOLEY, S. The effects of delayed reinforcement and post-feedback interval on learning in children. Unpublished master's thesis, University of Illinois, 1968.

AUTHOR INDEX

A

Abbott, Martha S., 203
André, J., 407

B

Birnbrauer, J. S., 440, 492

C

Campbell, Dugal, 282
Campione, Joseph C., 387
Cheyne, J. Allan, 231
Clifton, Rachel Keen, 122
Cohen, Leslie B., 419
Collins, Doris, 52
Cooper, Betty P., 203
Cunningham, Thomas F., 355

D

Danielson, Gary, 467
Davies, Graham M., 448
Dubignon, Judith, 282

F

Fixsen, Dean L., 96

G

Gelman, Rochel, 167
Ginsberg, Rose, 265
Girarbeau, Frederic L., 96
Gladstone, Roy, 31
Goldberg, Susan, 70
Gollin, Eugene S., 541
Goodman, Joseph, 553
Goulet, L. R., 434, 479

H

Haith, Marshall M., 52
Hall, Vernon C., 400
Hamilton, John, 114
Harris, Lauren, 374
Hill, Kennedy, T., 153
Hoyer, William J., 434

K

Karmel, Bernard Z., 339
Keeley, Stuart M., 578
Kessen, William, 52
Kolodny, May, 21

L

Lewis, Michael, 70
Lubker, Bonnie J., 188

M

Meichenbaum, Donald, 553
Meyer, William J., 326
Meyers, William J., 122
Mitler, Merrill M., 374
Moely, Barbara E., 153
Myers, Jerome L., 255
Myers, Nancy A., 255

N

Nakamura, Charles Y., 514

O

Orgel, Arthur, 326

P

Paraskevopoulos, Ioannis, 195
Porter, Walter L., 485
Pritchard, Suzanne, 459
Pronko, N. H., 136

R

Rabinovitch, M. Sam, 1
Rapoport, Judith L., 366
Redd, William H., 440
Rovee, Carolyn Kent, 245

S

Saravo, Anne, 21, 541
Scott, Keith G., 143
Sheppard, W. C., 36
Shriner, Thomas H., 532
Sibley, Sally A., 203
Smith, Jerome, 274
Snow, Catherine E., 1
Spradlin, Joseph E., 96

SPRAGUE, ROBERT L., 532, 578
STANDAHL, JON, 114

T

TIGHE, LOUISE S., 566
TIGHE, THOMAS J., 566
TOISTER, RICHARD P., 492
TUNICK, JEFFERY, 274

U

URE, DOROTHY, 220

W

WAHLER, ROBERT G., 101
WALTERS, RICHARD H., 231
WATSON, JOHN S., 467
WEISBERG, PAUL, 81
WHITEHURST, GROVER J., 314
WILCOX, BARBARA MORGAN, 10
WILLOUGHBY, ROBERT H., 299

Y

YUILLE, JOHN C., 459

SUBJECT INDEX

A

Age trends
  concept-identification, 374
  discrimination learning, 314, 407, 434
  oddity learning, 541
  operant behavior, 553
  paired associate learning, 400, 459
  pattern preferences, 339
  speech perception, 532
  stimulus generalization, 326
Attention, infants, 419, 467

C

Concept identification, 374
Constancy, size, 366

D

Deafness, 407
Discrimination
  complexity effects, 314
  in deaf children, 407
  reversal, 492
  shift effects, 407
  size, 366
  transposition, 566
Discriminative learning, 314, 434, 479, 485, 492, 541, 566
  dimensional shifts, 355
  retardates, 355, 387
  retention, 355
  transitivity effects, 387

G

Generalization
  mediated, 326
  stimulus, 326

H

Habituation, in infants, 419
Heart-rate, neonatal, 282

I

Infants
  attentional processes, 467
  pattern preferences, 339
  visual perception, 419

L

Learning
  discriminative, 314, 434, 479, 485, 492, 541, 566
  free operant, 492
  oddity, 541
  paired-associate, 400, 459
  relational, 485
  verbal, 459
  set, 541

M

Mediation
  by imagery, 459
  verbal, 448, 459, 479, 541
Modality, learning differences, 400

N

Neonates, 282

O

Observing responses, in infants, 419
Operants
  in pre-school children, 299
  motor, 553
  reward effects, 514
  schedules, 492
  verbal, 553

P

Perception, 566
Preferences
  for stimulus dimensions, 374
  in infants, 339, 419
Punishment
  by time-out from reinforcement, 299
  in discriminative learning, 314

R

Recognition, 448
Reinforcement
  delay, 578
  schedules, 514
  social, 440
  time-out effects, 299

Retardate
discriminative learning, 355, 387, 492, 578
reinforcement, 440
Retention, 448
Reward, magnitude, 514

S

Shift
optional, 407
reversal, 407
Speech, perception, 532
Stimuli, discriminative, 440
Stimulus, complexity, 339
Sucking, neonatal, 282

T

Temporal effects, time-out from reinforcement, 299
Transfer
in paired-associate learning, 400
shift behavior, 407
Transitivity, in discriminative learning, 387
Transposition, 566

V

Verbal learning, 434
modality differences, 400

JOURNAL OF Experimental Child Psychology

EDITOR: Sidney W. Bijou

ASSOCIATE EDITORS: Richard S. Bogartz
David Zeaman

EDITORIAL BOARD:

Donald M. Baer
Albert Bandura
Wesley C. Becker
Harry Beilin
Gordon N. Cantor
Peter Eimas
Robert L. Fantz
Eleanor J. Gibson
Eugene S. Gollin
Willard W. Hartup
Raymond H. Hohle
Herbert Kaye
Tracy S. Kendler
Helen L. Koch
Eugene R. Long
David S. Palermo
Herbert L. Pick
Hayne W. Reese
Bryan E. Shepp
Murray Sidman
Einar Siqueland
Patrick Suppes
Joachim Wohlwill

Volume 7, 1969

ACADEMIC PRESS
New York and London

Copyright © 1969, by Academic Press, Inc.

ALL RIGHTS RESERVED

No part of this volume may be reproduced in any form, by photostat, microfilm, or any other means, without written permission from the publishers.

CONTENTS OF VOLUME 7

Number 1, February 1969

Catherine E. Snow and M. Sam Rabinovitch. Conjunctive and Disjunctive Thinking in Children 1

Barbara Morgan Wilcox. Visual Preferences of Human Infants for Representations of the Human Face 10

Anne Saravo and May Kolodny. Learning Set and Shift Behavior in Children 21

Roy Gladstone. Age, Cognitive Control and Extinction . . . 31

W. C. Sheppard. Operant Control of Infant Vocal and Motor Behavior 36

Marshall M. Haith, William Kessen, and Doris Collins. Response of the Human Infant to Level of Complexity of Intermittent Visual Movement 52

Michael Lewis and Susan Goldberg. The Acquisition and Violation of Expectancy: An Experimental Paradigm 70

Paul Weisberg. Operant Procedures for the Establishment of Stimulus Control in Two-Year-Old Infants 81

Joseph E. Spradlin, Dean L. Fixsen, and Frederic L. Girarbeau. Reinstatement of an Operant Response by the Delivery of Reinforcement during Extinction 96

Robert G. Wahler. Infant Social Development: Some Experimental Analyses of an Infant-Mother Interaction during the First Year of Life 101

John Hamilton and Jon Standahl. Suppression of Stereotyped Screaming Behavior in a Profoundly Retarded Institutionalized Female 114

Rachel Keen Clifton and William J. Meyers. The Heart-Rate Response of Four-Month-Old Infants to Auditory Stimuli . 122

N. H. Pronko. Biotelemetry in Child Study 136

Keith G. Scott. The Design of Mobile Laboratories for Behavioral Research with Children 143

Kennedy T. Hill and Barbara E. Moely. Social Reinforcement as a Function of Task Instructions, Sex of *S*, Age of *S*, and Baseline Performance 153

Number 2, April 1969

Rochel Gelman. Conservation Acquisition: A Problem of Learning to Attend to Relevant Attributes 167

BONNIE J. LUBKER. Setting Similarity and Successive Discrimination Learning by Children 188
IOANNIS PARASKEVOPOULOS. Effects of Redundancy on Information-Reduction Tasks 195
SALLY A. SIBLEY, MARTHA S. ABBOTT, AND BETTY P. COOPER. Modification of the Classroom Behavior of a Disadvantaged Kindergarten Boy by Social Reinforcement and Isolation . . 203
DOROTHY URE. Spelling Performance of Left-Handed Schoolchildren as Affected by the Use of a Pencil Modified to Increase Visual Feedback 220
J. ALLAN CHEYNE AND RICHARD H. WALTERS. Intensity of Punishment, Timing of Punishment, and Cognitive Structure as Determinants of Response Inhibition 231
CAROLYN KENT ROVEE. Psychophysical Scaling of Olfactory Response to the Aliphatic Alcohols in Human Neonates . . . 245
NANCY A. MYERS AND JEROME L. MYERS. A Developmental Study of Sequence Structure in Binary Prediction 255
ROSE GINSBERG. Number and Color Responses in the Young Child 265
JEROME SMITH AND JEFFREY TUNICK. Cross-modal Transfer of a Discrimination by Retarded Children 274
JUDITH DUBIGNON AND DUGAL CAMPBELL. Sucking in the Newborn during a Feed 282
ROBERT H. WILLOUGHBY. The Effects of Time-Out from Positive Reinforcement on the Operant Behavior of Preschool Children 299
GROVER J. WHITEHURST. Discrimination Learning in Children as a Function of Reinforcement Condition, Task Complexity, and Chronological Age 314
WILLIAM J. MEYER AND ARTHUR ORGEL. The Effects of Age, Stimulus Intensity, and Training Trials on Mediated Stimulus Generalization 326
BERNARD Z. KARMEL. The Effect of Age, Complexity, and Amount of Contour on Pattern Preferences in Human Infants . . . 339
THOMAS F. CUNNINGHAM. Dimensional Transfer and Cue Retention in Retardate Discrimination Learning 355
JUDITH L. RAPOPORT. Size-Constancy in Children Measured by a Functional Size-Discrimination Task 366
MERRILL M. MITLER AND LAUREN HARRIS. Dimension Preference and Performance on a Series of Concept Identification Tasks in Kindergarten, First-Grade, and Third-Grade Children . . 374
ERRATA 385

NUMBER 3, JUNE 1969

JOSEPH C. CAMPIONE. Transitivity and Choice Behavior . . . 387
VERNON C. HALL. Acquisition and Transfer Differences between

Kindergarteners and Second-Graders on Aurally and Visually Presented Paired-Associates Using an A-B, A-C Design . . 400
J. André. Reversal Shift Behavior and Verbalization in Two Age Groups of Hearing and Deaf Children 407
Leslie B. Cohen. Observing Responses, Visual Preferences, and Habituation to Visual Stimuli in Infants 419
L. R. Goulet and William J. Hoyer. The Effects of Verbalization on Verbal Discrimination Learning and Associative Recall in Young Children and Adults 434
William H. Redd and J. S. Birnbrauer. Adults as Discriminative Stimuli for Different Reinforcement Contingencies with Retarded Children 440
Graham M. Davies. Recognition Memory for Pictured and Named Objects 448
John C. Yuille and Suzanne Pritchard. Noun Concreteness and Verbal Facilitation as Factors in Imaginal Mediation and PA Learning in Children 459
John S. Watson and Gary Danielson. An Attempt to Shape Bidimensional Attention in 24-Month-Old Infants 467
L. R. Goulet. The Effect of Verbalization on Discrimination Learning and Transfer in Nursery School Children 479
Walter L. Porter. Absolute versus Relational Responding in Children 485
Richard P. Toister and J. S. Birnbrauer. Facilitation of Repeated Free Operant Discrimination Reversals 492
Charles Y. Nakamura. Effects of Increasing and Decreasing Reward Magnitude and Pre-experimental Persistence Level on Focal and Incidental Responses 514
Thomas H. Shriner and Robert L. Sprague. Effects of Time-Compressed Speech Signals on Children's Identification Accuracy and Latency Measures 532
Anne Saravo and Eugene S. Gollin. Oddity Learning and Learning Sets in Children 541
Donald Meichenbaum and Joseph Goodman. The Developmental Control of Operant Motor Responding by Verbal Operants 553
Thomas J. Tighe and Louise S. Tighe. Perceptual Variables in the Transposition Behavior of Children 566
Stuart M. Keeley and Robert L. Sprague. Effects of Varying Delay of Reinforcement and Postreinforcement Intervals on Learning of Retardates 578

Author Index 585

Subject Index 587

A New Book from IASP

Contemporary Soviet Education

Readings from Soviet sources selected from **SOVIET EDUCATION**

Edited by Fred Ablin

Introduction by George S. Counts

This comprehensive collection of articles emphasizes the latest Soviet scholarship, practice and experimentation in the field of education, although there are also valuable materials on the development of education in the USSR since 1917.

The volume documents both positive and negative aspects: for example, the achievements of Soviet mathematics education, on the one hand, and the damage done to Soviet genetics education during the Stalin period, on the other.

Professor Counts' Introduction is a substantial evaluation of Soviet education as it has developed during the past fifty years.

Partial Table of Contents

The Spiritual Needs of Students in the Upper Grades: IU. V. SHAROV

Fifty Years of the Soviet Psychology of Learning: N. A. MENCHINSKAIA

The Drafts of the New Curricula for Grades 1-3 of the Elementary School and Their Testing

Draft of the Mathematics Program for Grades 1-3

Draft of the Secondary School Mathematics Program

On the Development of Genetics in Our Country and Scientific Truth: M. D. GOLUBOVSKII

Public Education in the USSR in Figures: M. POLUBOIARINOV

305 pp (8½" x 11") / LC 68-14428 $12.00

iasp

INTERNATIONAL ARTS AND SCIENCES PRESS
901 North Broadway, White Plains, New York 10603

Research in Verbal Behavior and Some Neurophysiological Implications

Edited by **Kurt Salzinger**
and **Suzanne Salzinger**
Biometrics Research
New York State Department of Mental Hygiene
and Polytechnic Institute of Brooklyn
New York, New York

This volume contains a complete discussion of the current directions of re search in language, including verbal behavior in the light of the principles o neurophysiology. It presents a study of language in man and animals, in it normal state and under the influence of drugs and pathology (aphasia, schizo phrenia) in adults and children, based upon models from the areas of learnin theory, linguistics, mathematics, psychophysics and neurophysiology. It include a discussion of problems emerging when techniques derived from animal ex perimentation are applied to human verbal behavior.

The work covers a wide range of topics, from vocal behavior in animals to emo tions and images in man. Fields discussed range from physiology to bilingua ism. Disciplines from linguistics and psychology to medicine, psychiatry psychopharmacology and neurology are represented.

The book will be particularly useful to graduate students and advanced under graduates taking courses in language as well as to research workers in the area

CONTRIBUTORS: Marianne Amarel. Louis Cassotta. Richard Allen Chase. Frances E Cheek. Charles Clifton, Jr. Alberto DiMascio. Stanley Feldstein. Samuel Fillenbaum Norman Geschwind. Norman Ginsburg. Murray Glanzer. Frieda Goldman-Eisler. Louis A. Gottschalk. Joseph H. Grosslight. Katherine Harris. Gilbert Honigfeld. Davis Howes Joseph Jaffe. Lyle V. Jones. Leonard Krasner. Wallace E. Lambert. Harlan L. Lane John C. Lilly. A. R. Luria. Paula Menyuk. Irwin Pollack. Malcolm S. Preston. Kurt Sal zinger. Arthur W. Staats. John A. Starkweather. Irene E. Waskow. Joseph M. Wepman Wesley C. Zaynor.

1967, 510 pp., $11.00

AP 2033

Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

A Study of Franz Brentano

His Psychological Standpoint and His Significance in the History of Psychology

By **Antos C. Rancurello**
Department of Psychology
University of Dayton, Dayton, Ohio

CONTENTS:

FRANZ BRENTANO: A PORTRAIT OF HIS LIFE, PERSONALITY AND WORK

OVERVIEW OF BRENTANO'S STANDPOINT IN PSYCHOLOGY

Introduction: Scientific Methodology and the Scope of Human Knowledge. Psychology from the Empirical Standpoint. Genetic or Explanatory Psychology. Descriptive or Phenomenological Psychology.

HISTORICAL PERSPECTIVE

The Emergence and Founding of Psychology (1825-1875): Brentano as an Interpreter of the Zeitgeist.

Early Growth of Psychology (1875-1900): Postscript on Brentano and Wundt and Their Relative Influence upon Psychology during This Period.

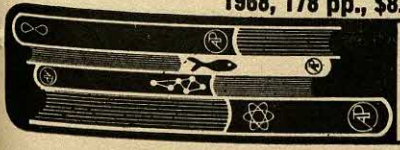
Psychology in Search for Identity (1900-1930): An Overview of Historical Trends and of Their Relationship to Brentano's Standpoint.

Progress of Psychology toward Maturity (1930-Present): Epilogue on Brentano, the "Third Force" Movement, and the Contemporary American Setting in Psychology.

ANNOTATED BIBLIOGRAPHY
AUTHOR INDEX—SUBJECT INDEX.

1968, 178 pp., $8.00

AP 2136



Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

AP

Experiments in Second-Language Learning

By **Edward Crothers**
and **Patrick Suppes**
Both at the Institute for Mathematical Studies in the Social Sciences
Stanford University, Stanford, California

This volume views second-language learning in the light of mathematical learning theory. It describes a series of sixteen experiments involving the acquisition of Russian by native American college students and dealing with phoneme discrimination, phoneme-grapheme correspondence, and the learning of vocabulary, morphology and syntax. The authors formulate mathematical learning models and apply them to the experimental data.

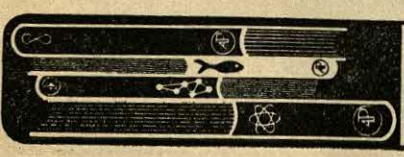
The book is written as a series of technical experimental reports. Discussions of psycholinguistic factors appear separately from the application of models.

CONTENTS:

Some Remarks on Stimulus-Response Theories of Language Learning
Learning to Discriminate Russian Phonemes
Learning the Orthographic Representations of Russian Sounds
Vocabulary Learning
Morphology Learning
Grammar Learning
Suggestions for Future Research
REFERENCES.
AUTHOR INDEX—SUBJECT INDEX.

1967, 374 pp., $13.50

AP 2024



Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

INFORMATION FOR AUTHORS

The *Journal of Experimental Child Psychology* will publish papers in which the behavior and development of children is clearly related to its determining variables. Typically (not always) this would mean that a variable has been manipulated in an experimental manner. Articles will be reviewed in terms of the apparent reliability of their findings and the clarity of the relationships they affirm. Research conceived in any theoretical framework is invited, so long as its conclusions are reliable and compelling. Occasionally, critical reviews of related bodies of experimental literature, theoretical discussions which define new paths of experimental work, and well-controlled developmental studies on infra-human subjects will be published. The criterion for acceptance of such papers will be their value in prompting experimental research in related areas with children. Short notes describing new apparatus or techniques, new materials, circuits, or other information valuable to investigators in the field will be accepted from time to time. There is no page charge for publication.

Two copies of each manuscript and each figure or table should be sent to:

Dr. Sidney W. Bijou, Child Behavior Laboratory, 403 E. Healey, University of Illinois, Champaign, Illinois 61820

A third copy should be retained by the author. Manuscripts must be double-spaced and may be typed, dittoed, or mimeographed. Each manuscript must include a concise abstract, preferably not exceeding 300 words, which will appear after the title and author's name in the published article. In some papers this abstract will obviate the need for a summary. In general, authors are free to use whatever organization their paper seems to require, but the specifications of the *Publication Manual* of the American Psychological Association are strongly recommended.

Figures and Tables. Figures should be numbered consecutively with Arabic numerals in order of mention in the text; each figure should have a descriptive legend. Legends should be typed together on a separate sheet, double-spaced. All illustrations should be in finished form ready for the engraver. A convenient size for drawings is $8\frac{1}{2} \times 11$ inches (21×27.5 cm). Drawings should be made with India ink on tracing linen, smooth surface white paper, or bristol board. Graph paper if used should be ruled in blue. Grid lines that are to show in the final engraving should be inked in black. Sharp, clear glossy prints of figures are acceptable.

Tables should be typed on separate pages, numbered consecutively with Arabic numerals in order of mention in the text. All tables should have descriptive headings, typed (double-spaced) above the table.

Literature references in the text should be cited by author's surname and the year of publication, e.g., Smith, 1960; Smith *et al.*, 1959; Smith and Jones, 1961, p. 250 (for references to a specific page). If a reference has more than two authors, the citation includes the surnames of all authors at the first mention, but later citations of the same reference include only the surname of the first author and the abbreviations *"et al."* Suffixes *a*, *b*, etc., should be used following the date to distinguish two or more works by the same author(s) in the same year, e.g., Smith, 1960a, 1960b. Literature citations in the bibliography should be arranged alphabetically according to the surname of the author.

Journal titles should be written out in full according to the form followed in the Psychological Abstracts (Vol. 39).

The desired style of capitalization and punctuation in references is apparent in the following examples:

STAATS, A. W., AND STAATS, CAROLYN K. A comparison of the development of speech and reading behavior with implications for research. *Child Development*, 1962, 33, 831–846.

SPIKER, C. C. Research methods in children's learning. In P. H. Mussen (Ed.), *Handbook for research methods in child development*. New York: Wiley, 1960. Pp. 374–420.

MUNN, N. L. *The evolution and growth of human behavior*. Cambridge: Riverside Press, 1955.

Reprints: Fifty gratis reprints are supplied and additional copies may be ordered.

Negro and White Children

A PSYCHOLOGICAL STUDY IN THE RURAL SOUTH

By **E. Earl Baughman**
W. Grant Dahlstrom
Department of Psychology, University of North Carolina Chapel Hill, North Carolina

1968, 572 pp., $15.00

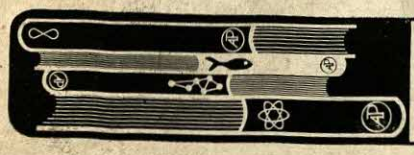
Reports the scientific conclusions of a unique four year study of racially-segregated students in a rural, economically deprived area of North Carolina.

CONTENTS:

SETTING AND METHODS
Life in Millfield
Procedures Used to Study Millfield's Children
INTELLIGENCE AND ACADEMIC ACHIEVEMENT
Ability Differences between Racial, Sex, and Age Groups
Academic Achievement Differences between Racial, Sex, Age, and Ability Groups
The Relationship of Selected Family Variables to Ability and Academic Achievement.
INTELLECTIVE CHANGE
Changes in IQ Among Young Negro Children
Changes in IQ Among White Adolescents
The Effect of Kindergarten Training upon the Intellective Functioning of Negro and White Children
A Pilot Project To Accelerate the Intellective Development of Four-Year-Old Negro and White Children
PERSONAL ATTRIBUTES
Millfield's Teachers View Their Pupils
Eighth-Grade Children as Assessed by the Minnesota Multiphasic Personality Inventory
Fantasy Behavior: Achievement Motivation
Fantasy Behavior: Other Attributes
Correlates of Peer Popularity
PERSONAL PERSPECTIVES
Mothers of Kindergarten-Age Children Report on Their Children, Training Practices, and Families
Eighth-Grade Children Report on Their Worlds
INTEGRATION AND RECOMMENDATIONS
Some Summary Comments and Observations
Narrowing the Gap

REFERENCES.
APPENDIX
AUTHOR INDEX—SUBJECT INDEX.

AP 2155



Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

JOURNAL OF Experimental Child Psychology

EDITOR: Sidney W. Bijou

ASSOCIATE EDITORS: Richard S. Bogartz
David Zeaman

EDITORIAL BOARD:

Donald M. Baer
Albert Bandura
Wesley C. Becker
Harry Beilin
Gordon N. Cantor
Peter Eimas
Robert L. Fantz
Eleanor J. Gibson
Eugene S. Gollin
Willard W. Hartup
Raymond H. Hohle
Herbert Kaye
Tracy S. Kendler
Helen L. Koch
Eugene R. Long
David S. Palermo
Herbert L. Pick
Hayne W. Reese
Bryan E. Shepp
Murray Sidman
Einar Siqueland
Patrick Suppes
Joachim Wohlwill

Volume 8, Number 1, August 1969

ACADEMIC PRESS
New York and London

CONTENTS

ALBERT KOZMA. The Effects of Anxiety, Stimulation, and Isolation on Social Reinforcer Effectiveness 1

JOSEPH F. FAGAN III. Free Recall Learning in Normal and Retarded Children 9

BRUCE M. ROSS AND JAMES YOUNISS. Ordering of Nonverbal Items in Children's Recognition Memory 20

CAROLYN KENT ROVEE AND DAVID T. ROVEE. Conjugate Reinforcement of Infant Exploratory Behavior 33

R. E. HERRON AND M. J. FROBISH. Computer Analysis and Display of Movement Patterns 40

DONALD E. P. SMITH, DALE BRETHOWER, AND RAYMOND CABOT. Increasing Task Behavior in a Language Arts Program by Providing Reinforcement . 45

FRANCIS J. DI VESTA AND MARTIN E. BERNSTEIN. The Effect of Serial Structure on Children's and Adults' Paired-Associate Learning 63

GORDON N. CANTOR AND SUNNAN K. KUBOSE. Preschool Children's Ratings of Familiarized and Nonfamiliarized Visual Stimuli 74

ERNEST A. LUMSDEN AND J. KENNETH KLING. The Relevance of an Adequate Concept of "Bigger" for Investigations of Size Conservation: A Methodological Critique 82

PETER WATSON. Partial Reward and Double Alternation Learning in Children 92

FRANK D. MILLER. The Effects of Differential Amounts of Stimulus Familiarization on Choice Reaction Time Performance in Children 106

GREGORY T. FOUTS AND DAVID A. PARTON. Imitation: Effects of Movement and Static Events 118

J. ALLAN CHEYNE, JOHN R. M. GOYECHE, AND RICHARD H. WALTERS. Attention, Anxiety, and Rules in Resintance-To-Deviation in Children . . 127

RONALD GALLIMORE, ROLAND G. THARP, AND BRYAN KEMP. Positive Reinforcing Function of "Negative Attention" 140

R. KEITH VAN WAGENEN, LEE MEYERSON, NANCY J. KERR, AND KURT MAHONEY. Field Trials of a New Procedure for Toilet Training . . 147

PETER D. EIMAS. A Developmental Study of Hypothesis Behavior and Focusing 160

ADRIENNE MOFFETT. Stimulus Complexity as a Determinant of Visual Attention in Infants 173

ANN M. CLARKE, G. M. COOPER, AND ELIZABETH H. LOUDON. A Set to Establish Equivalence Relations in Pre-school Children 180

DONALD H. KAUSLER AND MARY ANN GOTWAY. R-S Learning in Children . 190

Published bimonthly at Exchange Place, Hanover, Pa. 17331 by Academic Press, Inc., 111 Fifth Avenue, New York, N. Y., 10003.

In 1969, Volumes 7 and 8 (consisting of 3 issues each) will be published.

Price: $22.00

Private subscriptions (for subscriber's personal use): $10.00

All correspondence and subscription orders should be sent to the office of the Publishers at 111 Fifth Ave., New York, N.Y. 10003

Send notices of change of address to the office of the Publishers at least 4 weeks in advance.

Please include both old and new addresses.

Second-class postage paid at Hanover, Pennsylvania 17331

© 1969 by Academic Press, Inc.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 1–8 (1969)

The Effects of Anxiety, Stimulation, and Isolation on Social Reinforcer Effectiveness[1]

ALBERT KOZMA
Memorial University of Newfoundland

The present study investigated the effects of social isolation on subsequent social reinforcer effectiveness. High and low anxious Grade-3 *Ss* underwent 0, 3, 6, 12, and 18 minutes of isolation under stimulation and nonstimulation conditions. Following isolation, *Ss* received social reinforcement for correct responses on a probability matching task.

A linear relationship was obtained between length of isolation and reinforcer effectiveness for low anxious *Ss*. High anxious *Ss* showed increased susceptability to social reinforcement only after brief and prolonged periods of isolation. Stimulation appeared to delay susceptability to social reinforcement for both low and high anxious *Ss*.

Social isolation studies have consistently demonstrated that a brief period of isolation leads to increased social reinforcer effectiveness on a subsequent learning task (Hill and Stevenson, 1964; Lewis and Richman, 1964; Lewis, 1965). Despite these consistencies in the data, explanations for the findings have differed. The only point of agreement among researchers has been the postulation of an isolation-induced drive. Gewirtz and Baer (1958a,b), operating within a satiation-deprivation paradigm, assumed that the reinforcing value of a stimulus varies directly with its prior availability. They stipulated that the aroused drive results from a lack of "social" stimulation during isolation, and depends on social reinforcement for its reduction. As isolation normally also leads to a reduction in "sensory" or general stimulation, Stevenson and Odom (1962) argued that the drive is induced through sensory deprivation and could be reduced by any stimulus, including a social one.

An unanticipated consequence of isolation appeared to be an increase in *Ss*' anxiety. Walters and Ray (1960) claimed that such a "nonspecific state of arousal" leads to a general increase in drive, the reduction of which does not require a social reinforcer. The confounding of "deprivation operations" and "anxiety" during isolation led to the development

[1] This study is based on a thesis submitted by the author to the Department of Psychology, University of Western Ontario, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

© 1969 by Academic Press, Inc.

of a nonisolation technique for demonstrating the existence of social drives (Gewirtz, 1968). Although the new approach appears to have great relevance on the social drive issue, it seems entirely irrelevant to the development of an adequate explanation for the isolation data.

Recent findings (Dorwart *et al.*, 1965; Erickson, 1962) that social reinforcers are more effective in increasing performance on a learning task than nonsocial ones following isolation suggest that both Stevenson's and Walters' positions need modification. Moreover, Lewis' (1965) data make it difficult to retain the Gewirtz and Baer explanation. Lewis obtained a social isolation effect (increased social reinforcer effectiveness following social isolation) for 3- and 12-minute isolation intervals, but failed to find one for 6- and 9-minute intervals. These results point to a nonlinear relationship between isolation length and social reinforcer effectiveness, whereas the Gewirtz and Baer explanation would seem to require a linear relationship.

Lewis invoked an anxiety explanation for his findings. He suggested that initial (3 minute) isolation induced anxiety "over what was going to occur when the unfamiliar person (*E*) . . . returned"; that with the passing of time, interest in the environment led to visual exploration and a reduction in anxiety (6- and 9-minute intervals); that anxiety returned once the visual field was exhausted (by 12 minutes).

The preceding explanation of the social isolation effect (SIE) attributes the development of anxiety to a specific source and will therefore be referred to as the *specific anxiety* hypothesis.

Since in the explanation anxiety presumably develops because of an anticipated evaluation by E in a test-like situation, a positive evaluation on the subsequent task might be expected to reduce such anxiety more than other types of reward. If this assumption is accepted, the specific anxiety hypothesis would escape the difficulties posed for alternate explanations of the SIE by Erickson (1962) and Lewis (1965) studies.

The specific anxiety explanation leads to the following two major predictions: (a) An increase in visual stimulation during isolation will postpone the recurrence of the SIE, and (b) a longer isolation interval will be required for low test anxious (LA) *S*s than for high test anxious (HA) *S*s to produce a SIE. The first prediction is derived from the competing response role attributed to visual stimulation. Visual exploration supposedly reduces the anxiety produced by a brief isolation (3 minutes) period. Recurrence of anxiety would be delayed until the visual field had been exhausted. The richer the visual field, the greater would be the length of exploration and more prolonged the recurrence of anxiety. Since the SIE is attributed to anxiety aroused by isolation, its recurrence will also be delayed.

The second prediction follows from the assumption that *Ss* become anxious during isolation because they fear evaluation by a strange E in a test-like situation. HA *Ss* would tend to fear such evaluation more than LA *Ss* and should, accordingly, require less isolation time before becoming sufficiently anxious to show a SIE.

The present study was designed to test these two predictions.

METHOD

Subjects

Ss were 100 boys and 100 girls attending Grade 3 in London, Ontario, public schools.

Apparatus and Materials

A probability matching task was used to assess social reinforcer effectiveness. Two decks of 50 cards depicting cats and dogs in a 35:15 and 15:35 cat-dog ratio served as stimulus material on the learning task. This ratio was maintained for each set of ten cards. The order in which the two types of cards appeared was randomly determined for every ten cards with the exception that neither card could occur more than five consecutive times. Cards were dispensed from a plastic card dealer, and both cards and card dealer were kept in a box 18 inches long, 12 inches high and 6 inches deep, with an open side facing *E*.

Visual stimulation (during some isolation conditions) was provided by ten posters measuring 18 × 24 inches. All posters had vivid coloring and their content included food, furniture, appliances, cars, space-ships, and abstracts. In addition to the posters, the isolation room contained two fluorescent lights, a table, two chairs, and a one-way mirror.

Procedure

Median scores on the Test Anxiety Scale for Children were used to differentiate between HA and LA boys and girls. From the resulting four sex-anxiety groups, half the *Ss* were randomly assigned to one of two stimulation conditions in which the posters were either present (V) or absent (NV) during isolation. An equal number of *Ss* from the eight sex-anxiety-stimulation groups were randomly assigned to each of five social isolation conditions (0-, 3-, 6-, 12-, or 18-minute intervals).

Each *S* was taken individually to the isolation room by the male *E*, was seated and instructed to wait while *E* made some preparations. Following these instructions, *E* immediately entered the adjacent room and observed *S* for the appropriate isolation period. The 0 group was actually isolated for 1 second in order to keep instructions constant for all goups.

Upon the termination of the isolation period, *E* re-entered the room with the learning task. *S* was shown one card depicting a dog and one depicting a cat, was asked to identify the animals, and to state his preference. He was then told that he would be required to play a game in which he must guess which picture would apppear on the next trial. After *S* made his choice, he was presented with the correct event, whereupon the card was removed. All *S*s received social reinforcement for correct choices by *E*'s warmly saying either "good," "right," or "fine." Each *S*s was tested on the deck containing the higher ratio of his nonpreferred choice. The deck was administered twice, once forward, and once backward for a total of 100 trials. Responses were recorded on a check list.

At the termination of the learning task, *S* was given a brief questionnaire about his activities during isolation.

RESULTS

Learning scores (total number of choices of the more frequent event) were analyzed by a 2 (anxiety) $\times$ 2 (stimulation $\times$ 5 (isolation) analysis of variance design to assess the effects of anxiety, stimulation, and isolation length on social reinforcer effectiveness. Significant effects were obtained for stimulation ($p < .05$), isolation ($p < .01$), and the Anxiety $\times$ Stimulation $\times$ Isolation interaction ($p < .05$) (Table 1).

A plotting of the three-way interaction suggested increased social reinforcer effectiveness for the following groups: 3 HA NV, 12 HA NV, and 18 HA NV; 6 HA V; and 18 HA V; 6 LA NV, 12 LA NV, and 18 LA NV; 12 LA V and 18 LA V (Fig. 1).

In order to assess whether social reinforcer effectiveness was in fact greater for the groups suggested by the interaction, anxiety-stimulation

TABLE 1

ANALYSIS OF LEARNING SCORES FOR ANXIETY, STIMULATION, AND ISOLATION CONDITIONS

Source of variation	*df*	*MS*	*F*
Anxiety (A)	1	0.23	
Stimulation (B)	1	34.23	4.20*
Isolation (C)	4	30.31	3.72**
AB	1	0.00	
AC	4	4.22	
BC	4	7.52	
ABC	4	22.48	2.76*
SS within	180	8.16	

* $p < .05$.
** $p < .05$.

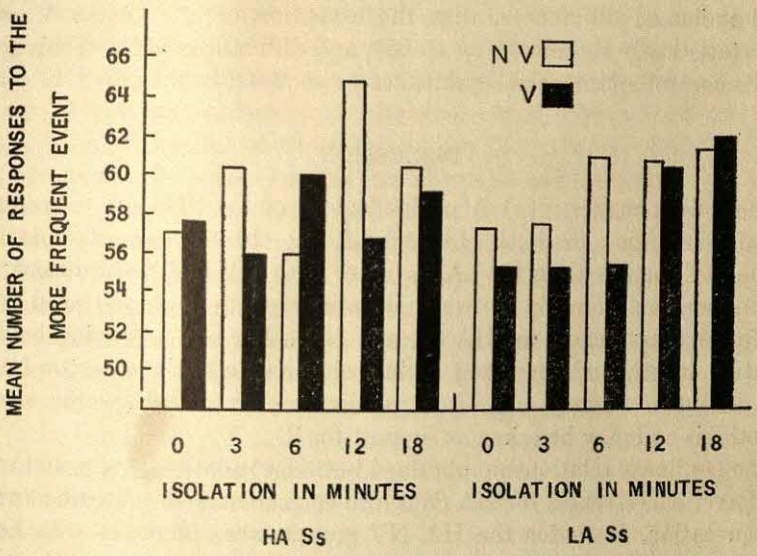


FIG. 1. A graphic illustration of the relationship among anxiety, stimulation, and isolation conditions.

groups at all isolation intervals were compared with their 0 control groups (Winer, 1962; p. 238). Since stimulation could not play a significant role for the 0 isolation groups (there was no time to look) Group 0 HA NV was combined with 0 HA V, and Group 0 LA NV was combined with 0 LA V for the comparison. The procedure gave a better estimate of the control means by doubling the number of *Ss* in each control group.

TABLE 2

DIFFERENCES IN LEARNING SCORES BETWEEN THE VARIOUS ISOLATION GROUPS AND THEIR 0 ISOLATION CONTROL UNDER ANXIETY AND STIMULATION CONDITIONS

		Isolation in minutes							
		3		6		12		18	
Anx.	Stim.	*D*	*F*	*D*	*F*	*D*	*F*	*D*	*F*
HA	NV	31	5.88*	−9	0.49	79	38.24**	31	5.89*
	V	−14	1.20	28	4.80*	13	1.04	18	1.99
LA	NV	11	0.74	47	13.54**	47	13.54**	53	17.21**
	V	−11	0.74	−8	0.39	43	11.33**	59	21.45**

*$p < .05$.
**$p < .01$.

All indicated differences, with the exception of the 18 HA V group were statistically significant ($p < .05$), and differences between this group and its control approached significance ($p < .20$) (Table 2).

DISCUSSION

These data suggest: (a) A postponement of the SIE due to increased visual stimulation for both HA and LA *Ss*; (b) the earlier appearance of the SIE for HA than for LA *Ss* under both NV and V conditions; (c) a nonlinear relationship between isolation length and increased social reinforcer effectiveness for HA *Ss*; and (d) a linear relationship between isolation length and increased social reinforcer effectiveness for LA *Ss*. The first two findings are direct predictions from the specific anxiety hypothesis and may be taken as support for it.

The nonlinear relationship obtained between isolation length and social reinforcer effectiveness for HA *Ss* is also consistent with a specific anxiety interpretation. Data for the HA NV groups were identical with Lewis' (1965) results, and his explanation may be invoked to account for them. It is suggested that under conditions of minimal stimulation during isolation *Ss* initially became anxious over an anticipated evaluation by a strange E,[2] that subsequent visual exploration temporarily reduced the anxiety, and that anxiety recurred once the visual field had been exhausted. Increased social reinforcer effectiveness at three, 12 and 18 minutes of isolation is attributed to peaks in anxiety at these periods for HA NV *Ss*.

The introduction of highly vivid stimulus material during isolation (HA V groups) might be expected to have led to cursory exploration, a delay in the development of anxiety, and a postponement of the SIE. After the termination of the brief exploration interval, anxiety would have occurred, only to become inhibited again once a more thorough examination began. Anxiety would have recurred with the exhaustion of the visual field. The SIE for the 6-minute HA V group can be attributed to a peak in anxiety at this isolation interval, and the approaching effect for the 18-minute HA V group suggests that these *Ss* had not entirely exhausted the more complex visual field.

Although the linear relationship obtained between isolation length and social reinforcer effectiveness for LA *Ss* was not predicted, these data may also be used to support a specific anxiety explanation of the SIE. By definition, LA *Ss* would have been less concerned then HA *Ss* after initial isolation, and would have required a longer incubation period to become sufficiently anxious to produce the SIE. During the longer period,

[2] Paivio's incubation phenomenon (Paivio, 1963).

they might be expected to have indulged in visual exploration and exhausted the visual field. In the absence of further exploration, anxiety would have remained once it occurred. If the SIE is attributed to peaks in anxiety at various isolation intervals, there should have been an initial delay in the occurrence of the effect for LA *Ss*; once the effect occurred, however, it would have remained for longer isolation intervals.

Increased visual stimulation during isolation for LA *Ss* would have led to a further postponement of the SIE. A relatively long isolation period would have been required for LA *Ss* to exhaust the more complex visual field and to bring about the onset of anxiety.

The explanation employed here to account for the findings on LA *Ss* calls for: (a) A longer isolation period for LA than HA *Ss* for the occurrence of the SIE; (b) a linear relationship between isolation length and social reinforcer effectiveness for LA *Ss*; and (c) a further postponement of the SIE for LA *Ss* if isolation occurs in a visually stimulating environment. The data on LA *Ss* in the present study fit the explanation perfectly.

REFERENCES

Dorwart, W., Ezerman, R., Lewis, M., and Rosenhan, D. The effects of brief social deprivation on social and non-social reinforcement. *Journal of Personality and Social Psychology*, 1965, **2**, 111–115.

Erickson, M. T. Effects of social deprivation and satiation on verbal conditioning in children. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 953–957.

Gewirtz, J. L. Deprivation and satiation of social stimuli as determinants of their reinforcing efficacy, in J. P. Hill, ed., *Minnesota Symposia on Child Psychology*, Vol. I., pp. 3–56. University of Minnesota Press, Minneapolis, 1967.

Gewirtz, J. L., and Baer, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *Journal of Abnormal and Social Psychology*, 1958, **56**, 49–56 (a).

Gewirtz, J. L., and Baer, D. M. Deprivation and satiation of social reinforcers as drive conditions. *Journal of Abnormal and Social Psychology*, 1958, **57**, 165–172 (b).

Hill, K. T., and Stevenson, H. W. The effectiveness of social reinforcement following social and sensory deprivation. *Journal of Abnormal and Social Psychology*, 1964, **68**, 579–584.

Lewis, M. Social isolation: A parametric study of its effects on social reinforcement. *Journal of Experimental Child Psychology*, 1965, **2**, 205–218.

Lewis, M., and Richman, S. Social encounters and their effects on subsequent reinforcement. *Journal of Abnormal and Social Psychology*, 1964, **69**, 253–257.

Paivio, A. U. Audience influence, social isolation and speech. *Journal of Abnormal and Social Psychology*, 1963, **67**, 247–253.

Stevenson, H. W., and Odom, R. D. The effectiveness of social reinforcement following two conditions of social isolation. *Journal of Abnormal and Social Psychology*, 1961, **63**, 147–154.

WALTERS, R. H., AND RAY, E. Anxiety, isolation, and reinforcer effectiveness. *Journal of Personality*, 1960, 28, 257–267.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 9–19 (1969)

Free Recall Learning in Normal and Retarded Children[1]

JOSEPH F. FAGAN III
Case Western Reserve University

General and operational distinctions were drawn between item and order properties and between primary and secondary organization in multitrial free recall (MFR) to explore the assumption that poor immediate memory on the part of retarded children is due to a deficiency in ordering. Twelve educable retardates and 12 normals, matched with the retardates for MA, were each given 10 free recall trials on a list of 10 letters drawn randomly from the alphabet. Estimates of item recall, primary and secondary ordering, and the frequency of omission, repetition, and intrusion errors were obtained. The results indicated (a) a retardate deficit in item recall, (b) no normal-retardate differences in primary or secondary ordering, and (c) a retardate susceptibility to errors of intrusion. These findings offer no support for the hypothesis that the retardate's lowered item recall is due to inefficient organization.

The available evidence suggests that retardates are inferior in immediate memory when compared to normal children of similar MA (Hermelin and O'Connor, 1964; Holden, 1965; Neufeldt, 1966; Fagan, 1966, 1968a). Spitz (1966) has proposed that the retarded child's lowered immediate recall is due to a deficiency in organization. His view is that the acquisition of information is determined by the organization, both imposed and intrinsic, of the incoming stimulation. Retardates are assumed to be less efficient than normals in imposing order on incoming material. Thus, the retardates takes in less information than the normal. While the ongoing organization of incoming material is difficult if not impossible to operationalize, the effects of such organization may be inferred to some degree from the order in which children retrieve items. To test Spitz's hypothesis, modes of ordering stimulation in a given experimental setting must be identified and measured. Some tasks are better suited for this purpose than others. The procedure usually employed in short-term memory research is to present a stimulus complex to *S* followed by an

[1]The author is indebted to Mr. Robert Evans, Principal of the Anderson School, Lyndhurst; Mr. George Baily, Principal of the Willowick Jr. High School; and Dr. Jane Kessler, Director of the CWRU Mental Development Center for providing subjects and facilities. The assistance of Mrs. Mary M. Parker in preparing materials and aiding in the scoring and analysis of data is gratefully acknowledged.

occasion for response and, when multiple trials are to be given, to introduce a new set of items on each trial. Such a situation contains few possibilities for the measurement of ordered recall, although attempts have been made in this direction (Fagan, 1968a). Multitrial free recall (MFR), on the other hand, is a fruitful paradigm for the observation of ordering. In MFR different arrangements of the same items are presented from trial to trial and *S* is allowed to recall the items in any order he wishes.

Bousfield and Bousfield (1966) have proposed that the analysis of MFR be directed toward item and order properties of the data. Item properties are indexed by sheer amount of material retained, while order properties are estimated from the sequence in which the items are recalled. Tulving (1968, pp. 15–16) further distinguishes between primary and secondary organization in MFR. According to Tulving, primary organization refers to "consistent discrepancies between input and output orders that are independent of the subject's prior familiarity with a set of items." The tendency to recall terminally presented items first (the recency effect) is cited as an example of primary organization. In the present study, primary organization was defined as consistent recall ordering based on the serial position of the item in the list. The specific index employed was the serial-position curve which reflects both primacy and recency strategies.

Secondary organization, for Tulving, is any consistent discrepancy between list and recall order which is determined by *S*'s "prior . . . acquaintance with the items constituting a list." Two approaches may be taken to the measurement of secondary organization. The first relies on *E*'s knowledge of sources of order in the list, such as interitem or categorical associations. The second is independent of *E*'s preconceptions and refers to *S*'s consistency in response ordering over a series of trials when different permutations of the same items are presented from trial to trial. Tulving (1962) uses the term subjective organization (SO) to describe this consistency, while Bousfield and Bousfield refer to the phenomenon as intertrial repetition (ITR). Since this second general approach to the measurement of secondary organization permits *S* to supply the criterion of order, it would seem to provide a most objective means of studying secondary organizational processes. Prior comparisons, however, of retarded and normal children of similar MA on MFR have relied exclusively on category clustering as an estimate of secondary ordering (Osborn, 1960; Rossi, 1963; Spitz, 1966). In the present study, secondary organization was indexed by a scoring system (Fagan, 1968b) based on the Bousfield and Bousfield ITR measure, a measure dependent on the child's criterion of order.

A further aspect of MFR data which may increase our knowledge of immature memory processes includes the types of errors which children commit. Conrad (1959, 1964) and Wickelgren (1965, 1966) have demonstrated the heuristic value of error analyses in short-term memory, serial recall experiments, while Craik (1968) has attempted to extend such analyses to MFR. The present experiment included a comparison of normals and retardates on the relative frequency of their omission, repetition, and intrusion errors.

In summary, the purpose of the present study was to compare retarded and normal children of similar MA on the item recall, primary and secondary ordering, and error characteristics of their MFR performance.

METHOD

Subjects

The sample consisted of 24 *S*s, 12 in each of the two groups. The retardates were educable and noninstitutionalized with no apparent speech, hearing, or emotional difficulties. They were drawn from suburban classrooms where the sex distribution happened to be nine males and three females for the CA, MA, and IQ restrictions required in this study. The normal *S*s were matched with the retardates for sex and MA. Relevant CA, MA, and IQ characteristics of the two groups are listed in Table 1. The CA and MA data are represented in years.

TABLE 1
CA, MA, AND IQ CHARACTERISTICS FOR EXPERIMENTAL GROUPS
(N = 12 PER GROUP)

Group	Measure	CA (years)	MA (years)	IQ
Normal	*M*	9.62	9.77	101.50
	SD	.60	.63	4.52
Retarded	*M*	13.02	9.80	75.25
	SD	.73	.49	3.69

Design

There were two groups: 12 normals (N) and 12 retardates (R). The task consisted of orally presented sequences of single letters requiring oral short-term recall. Each *S* was given 10 free recall trials on a list of 10 letters. Lists were constructed by selecting 10 different letters randomly from the alphabet and deriving 10 random orderings of the items, with the restriction that obvious letter sequences were avoided. Eight lists were constructed in this manner. Within each group, four *S*s received the

same lists and trial sequences as one other member of the sample. Across groups, *S*s were matched for the lists they were to recall.

Procedure

The *S*s were tested individually in a quiet room. The equipment consisted of a Panasonic, Model RQ-156S, Tape Recorder. From a prepared tape, *S* heard the word, "ready," followed 2 seconds later by the first letter in the message. The letters were presented at the rate of one every 2 seconds. The word, "now," followed the last letter in the message by two seconds. The *S*'s task was to recall the letters when he heard "now." Following the "now" signal on each trial, *S* was allowed 30 seconds for recall followed by 10 more seconds until the next "ready" signal. The words "ready," "now," and the letters in each message were recorded by a female assistant at approximately the same loudness. At the beginning of each session, the volume at which the tape was to be played was adjusted to a comfortable level for each *S* and was maintained at that level throughout the session.

Prior to test, *S* was told, "This is a game to see how well kids remember letters. You are going to hear 'ready' and then some letters from the alphabet. When you hear the word 'now,' tell me what the letters were. You can remember the letters any way you want to. Then you will hear the same letters again but they will be mixed up a different way. The object of the game is to remember as many letters as you can each time." Once it was clear that *S* understood what was expected of him, the experiment proper began with: "Fine! Now we're going to listen to the tape recorder. Remember the object of the game is to tell me what the letters were when you hear 'now.' You can remember the letters any way you want to." No *S* had any difficulty in following the instructions. A typical session lasted about 15 minutes.

RESULTS

The major objectives in statistical analyses were to determine the effects of IQ on item recall, primary and secondary organization, and the distribution of different kinds of errors. From the results of previous short-term memory studies, normals were expected to be superior to retardates on simple item recall. The combination of Spitz's assumption that lowered recall is due to deficient organizational processes and Tulving's distinction between two types of ordering led to normal-retardate comparisons on measures of both primary and secondary organization. Finally, given the demonstrated value of error analyses in a variety of memory tasks, an attempt was made to relate intellectual deficit to qualitative distinctions in errors. Specifically, for each experimental group

(N, R) we determined the number of items correctly retrieved (item recall), response ordering based on the serial position of the item in the list (serial-position curves), consistent trial-to-trial recall ordering based on *S*-determined inter-item associations (modified ITR scores), and the frequency of omission, repetition, and intrusion errors.

Item Recall

Item recall was a measure of amount of material retained with no value being attached to recall order. According to this scheme, *S* received one point for each letter correctly retrieved. If, for example, *S* were presented with the letters Q-E-C-R-T-A-K-X-H-V and his reply were T-X-H-G-R-U, his item recall score would be four, since he had retained the letters T,X,H, and R. A critical aspect for quantifying item recall was to estimate scores were *S* to respond randomly. Bousfield and Bousfield have developed a formula to determine the expected value of chance item recall based on the number of items in the message, h, the number of non-repeated items in the reply, k, the number of correctly recalled items, c, and the size of the pool from which the items were drawn, w. The expected chance value of c is given by the formula $h \cdot k/w$. In the present study, w, equaled 26, the size of the alphabet. In the example given above, *S* replied with two incorrect and four correct letters. Thus $h = 10$, $k = 6$, and $w = 26$, and the expected number correct by chance equals $10 \cdot 6/26 = 2.31$. Substracting 2.31 from the observed number of correct letters, 4, gives a corrected item recall score of 1.69. Under this scoring system the maximum corrected item recall score on any trial was $10 - [(10 \cdot 10)/26]$ or 6.16.

For each *S* there were 10 corrected item recall scores, one for each trial. From these measures the group's mean corrected item recall score was derived for each trial. These mean group scores are illustrated in Fig. 1. The corrected item recall scores were entered into a 2 (Groups) $\times$ 10 (Trials), repeated measures ANOVA. The superior recall of the normals, apparent from Fig. 1, proved to be statistically significant ($F = 10.82$, df 1/22, $p < .01$), as was the general increase in recall over trials for both groups ($F = 2.76$, df 9/198, $p < .01$). There was no reliable Groups $\times$ Trials interaction. Thus, the item recall results confirmed the expectation of a retardate deficit in immediate memory.

Primary Organization

The main analysis of primary ordering was based on serial-position curves, derived by plotting frequency of item recall against the serial position of the item in the message. This procedure usually results in curves which are bow-shaped with the greatest recall occurring at initial

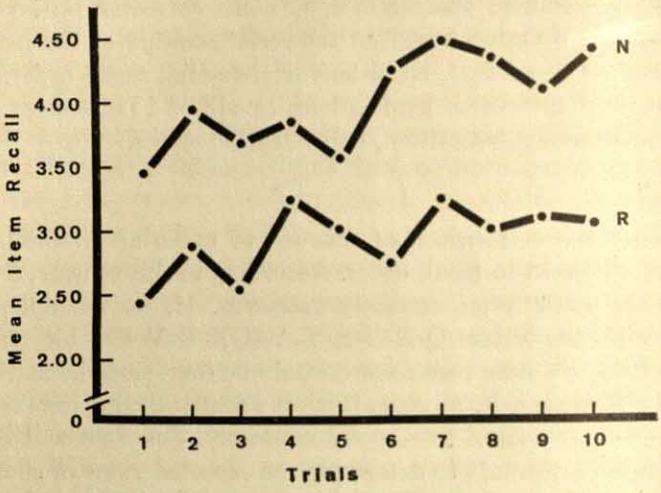


FIG. 1. Mean corrected item recall scores, over trials, for normal (N) and retarded (R) *Ss*.

and terminal serial positions. The latter phenomena are referred to as primacy and recency effects, respectively. In the present experiment, a discrepancy in amount or kind of primary ordering between normals and retardates would be reflected in differentially shaped serial-position curves. As one may see from Fig. 2, the curves for the normals (N) and the retardates (R) are virtually identical in form. Both are bowed with

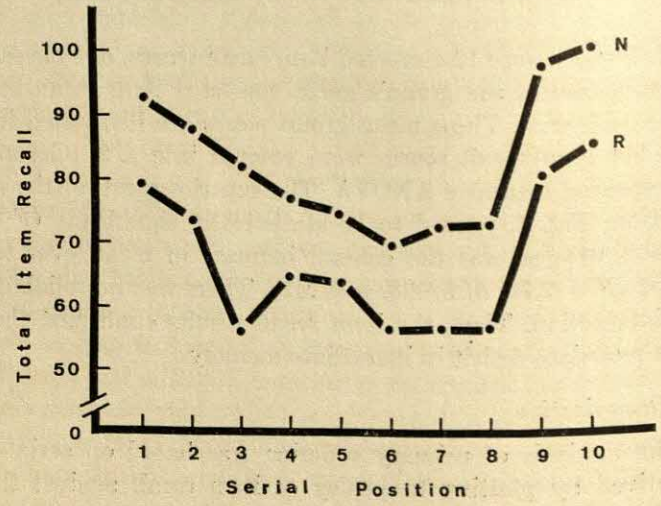


FIG. 2. Total item recall as a function of input serial position for normal (N) and retarded (R) *Ss*.

pronounced primacy and recency effects, and, as is usually the case in MFR, recall is somewhat greater for terminal than for initial items. The total number of items recalled at each serial position for each *S* in both the N and the R samples constituted the basic data for a 2 (Groups) × 10 (Serial Positions), repeated measures ANOVA. Main effects due to Groups and to Serial Position were highly significant ($F = 9.19$, df 1/22, $p < .01$ and $F = 10.29$, df 9/198, $p < .01$, respectively); the former indicating the superior performance of the normals on an uncorrected measure of item recall and the latter confirming the fact that recall depends on serial position. The Groups × Serial Position interaction, however, was statistically unreliable ($F < 1$), which means that differences in intelligence did not alter the form of the serial position curve. A secondary analysis, aimed at identifying particular primary ordering strategies, was undertaken by determining for each *S* on every trial his rank order of item recall in relation to the item's serial position in presentation. Two dominant ordering techniques emerged from this examination. Some children consistently began their recall with the first items presented. Five of the normals and four of the retardates employed such a "primacy" strategy. Other *S*s, four normals and five retardates, ordered their recall by retrieving the most recently presented items first. Three *S*s in each group exhibited no identifiable primacy or recency strategies. The results obtained from an inspection of both frequency and rank order of item recall in relation to input serial position lead to two conclusions: Recall ordering based on the serial position of the list item occurred but did not serve to differentiate normals from retardates.

Secondary Organization

Secondary organization was operationally defined as trial to trial sequential consistency (SC) in the forward ordering of correctly recalled items and was indexed by a modified version (Fagan, 1968b) of the Bousfield and Bousfield intertrial repetition (ITR) scoring system. The basic unit of consistency was the recurring bigram or ITR. The replies X-D-J-B and J-B-X-D-N, for example, contain the repeated bigrams X-D and J-B. SC was expressed as the ratio of the observed number of ITR's to the maximum possible, to obtain an estimate of amount of ordering for normals and retardates independent of differences in item recall. This observed measure of sequential consistency, O(SC), was determined by the formula $O(ITR)/c - 1$ where c is the number of items common to the two recalls. In the example given above, $O(ITR) = 2$, $c = 4$, and $O(SC) = .67$. To estimate the amount of SC expected by chance, E(SC), the formula $c/h \cdot k$ was employed, where h is the number of items recalled on trial t and k is the number of items recalled on trial $t + 1$. Thus, the

corrected SC score, O(SC) — E(SC), for our hypothetical *S* would be .67 — .20 or .47. Under this scoring system a maximum possible trial to trial O(SC) — E(SC) value was .90.

All *S*s received an O(SC), E(SC), and O(SC) — E(SC) score for each of the nine possible forward, adjacent, two-trial sequences 1–2, 2–3, . . . 9–10. The nine scores in each category were then summed to give *S* three total values. The mean total O(SC), E(SC), and O(SC) — E(SC) estimates for the N and R groups are listed in Table 2. Inspection of the

TABLE 2
MEAN TOTAL SC AND ERROR SCORES FOR THE N AND R SAMPLES

Score	Category	Measure	Group N	Group R
SC	O (SC)	M	1.60	1.31
		SD	.64	.81
	E (SC)	M	1.03	.99
		SD	.10	.10
	O (SC) – E (SC)	M	.57	.32
		SD	.66	.80
Errors	Om	M	17.83	21.33
		SD	12.06	12.42
	Rep	M	6.66	8.75
		SD	5.25	4.31
	Int	M	6.33	13.91
		SD	4.92	8.28

SC data in Table 2 elicits two impressions: The absolute magnitude of secondary ordering for each group appears to be quite low and minimal differences in SC between normals and retardates seem to be the case. There was some statistical confirmation for both of these inferences. The maximum possible mean total O(SC) — E(SC) score for each group was 8.10, since a value of .90 could have been achieved on each of the nine two-trial sequences. Group N's mean total O(SC) — E(SC) score of .57 and group R's of .32 represented only 7 and 4%, respectively, of the maximum possible. Correlated *t* tests were performed comparing O(SC) with E(SC) values within each sample to determine if greater than chance secondary ordering occurred. The normals' O(SC) was reliably above chance ($t = 3.00$, $df = 11$, $p < .01$), while the retardates' was not ($t = 1.39$, $df = 11$). Although this latter finding would seem to indicate a normal advantage in secondary organization, a direct comparison of normal with retardate O(SC) — E(SC) mean totals revealed no statistically significant differences in amount of ordering between groups ($t = .80$, $df = 22$). In short, the low absolute magnitude of SC for both samples

combined with the fact that corrected SC values for groups N and R were not reliably disparate makes it extremely doubtful that the obtained normal-retardate differences in item recall may be explained on the basis of a normal superiority in secondary organization.

Errors

On an MFR task, *Ss* may err by omission, repetition, or intrusion. In the present experiment, failure to respond *per se* was considered an error of omission (Om), the retrieval of an item not in the message was defined as an intrusion (Int), and the reiteration of either a correct or intruded response constituted a repetition (Rep). If retardates are no more prone to commit a certain type of mistake than normals, the relative frequencies of Om, Rep, and Int errors should be similar for the two groups. An unequal distribution of faults between samples, on the other hand, would suggest an IQ-related susceptibility to a particular variety of error. To assess this possibility, each *S*'s Om, Rep, and Int errors were summed over ten trials. The mean total errors in each category for groups N and R are presented in Table 2. An examination of the error data in Table 2 indicates that the frequency of intrusions constituted the most outstanding discrepancy in performance between the normals and the retardates. Group R imported double the number of nonlist items than group N. The slight disparities between samples of the amount of Om or Rep faults were not statistically reliable ($t = .68$, $df = 22$ and $t = .99$, $df = 22$, for Om and Rep errors respectively). The retardates, however, committed significantly more Int errors than the normals ($t = 2.65$, $df = 22$, $p < .02$). From these results we may infer that the retardates' tendency to retrieve non-list items contributed strongly to the item recall differences between groups.

DISCUSSION

The most important aspects of the present results lie in their implications for the "organizational deficit" hypothesis of Spitz (1966). To explore Spitz's assumption that lowered immediate item recall on the part of retarded *Ss* is due to a deficiency in ordering incoming material, general and operational distinctions were drawn between item and order properties and between primary and secondary retrieval organization in MFR. The pertinent results may be summarized as follows. The present finding that retarded children are deficient in immediate item recall on an MFR task when matched with equal-MA normals is concordant with the results of previous normal-retardate comparisons employing single-trial serial recall tests (Hermelin and O'Connor, 1964; Holden, 1965; Neufeldt, 1966; Fagan, 1966, 1968a). Measures of both primary and secondary

organization, however, did not serve to differentiate normal from retarded children. While these findings do not add credence to the "organizational deficit" hypothesis neither do they refute it, since theoretical and operational definitions of organization in MFR other than those employed here are possible. In any case, the present findings lend no support to the assumption that the retardate's poor item recall is due to inefficient ordering.

The high number of intrusion errors on the part of the retarded children in this study merits further comment. While a statistically satisfactory search for the source of the intrusions was precluded by the design of the experiment, systematic perusal of individual protocals did elicit some general impressions. In most cases, *Ss* were quite consistent in omitting and importing the same letters from trial to trial. Intrusions were usually acoustically similar to omitted items. If an M were forgotten, for example, an N was imported. Thus, it is possible that interference due to acoustic similarily may have played a part in the retardate's inferior performance. It is also possible that acoustic confusion may have occurred at two points in time: During input or at recall. In the first case, the item may have been simply misperceived; in the second, the letter may have been acquired but an acoustically similar item retrieved. Whether retarded children are particularly susceptible to acoustic confusion in the MFR of letters and whether such interference occurs during input or upon retrieval are empirical questions, explorations of which are currently in progress.

REFERENCES

Bousfield, A. K., and Bousfield, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, 1966, **19**, 935–942.

Conrad, R. Errors of immediate memory. *British Journal of Psychology*, 1959, **50**, 349–359.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–83.

Craik, F. I. M. Types of error in free recall. *Psychonomic Science*, 1968, **10**, 353–354.

Fagan, J. F. Short-term retention in normal and retarded children. *Psychonomic Science*, 1966, **6**, 303–304.

Fagan, J. F. Short-term memory processes in normal and retarded children. *Journal of Experimental Child Psychology*, 1968, **6**, 279–296. (a)

Fagan, J. F. Measuring verbal recall. II: the ITR score expressed as a ratio. *Psychonomic Science*, 1968, **11**, 205. (b)

Hermelin, B., and O'Connor, N. Short-term memory in normal and subnormal children. *American Journal of Mental Deficiency*, 1964, **69**, 121–125.

Holden, E. A. Temporal factors and subnormality in visual pattern recognition: a test of stimulus trace theory. *Journal of Comparative and Physiological Psychology*, 1965, **59**, 340–344.

NEUFELDT, A. H. Short-term memory in the mentally retarded: an application of the dichotic listening technique. *Psychological Monographs,* 1966, **80,** 1–31.

OSBORN, W. J. Associative clustering in organic and familial retardates. *American Journal of Mental Deficiency,* 1960, **65,** 351–357.

ROSSI, E. L. Associative clustering in normal and retarded children. *American Journal of Mental Deficiency,* 1963, **67,** 691–699.

SPITZ, H. H. The role of input organization in the learning and memory of mental retardates. In N. R. Ellis (Ed.), *International review of research in mental retardation* Vol. 2. New York: Academic Press, 1966.

TULVING, E. Subjective organization in the free recall of "unrelated" words. *Psychological Review,* 1962, **69,** 344–354.

TULVING, E. Theoretical issues in free recall. In T. R. Dixon and D. L. Horton Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs: Prentice-Hall, 1968.

WICKELGREN, W. A. Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology,* 1965, **70,** 102–108.

WICKELGREN, W. A. Phonemic similarity and interference in short-term memory for single letters. *Journal of Experimental Psychology,* 1966, **71,** 396–404.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 20–32 (1969)

Ordering of Nonverbal Items in Children's Recognition Memory[1]

BRUCE M. ROSS AND JAMES YOUNISS
Center for Research in Thinking and Language, Catholic University of America

In two experiments children of different ages were instructed to recognize certain items previously pointed out from nine simultaneously exposed nonverbal items. In Exp. 1, recognition conditions of two, three, or two out of three items were administered; in Exp. 2, children had to recognize three items either in the exact order shown or in any order. In both experiments type of material was a variable. The youngest children, CA 5 to 6, showed a significant amount of spontaneous ordering in easy conditions, while ordering in the hardest condition was low for these *Ss* but increased with age. A major result was that the oldest *Ss*, CA 10, recognized as many items correctly when they were requested to order items as when they were not; at CA 6 this finding held only for easy items. It is concluded that storage of order information is relatively independent from retention of item information in recognition memory.

In what order do individuals reinstate retained memory items? For verbal items the answer is quite clear; connected discourse is given back in forward serial order and unrelated words are clustered according to imposed meaning when free recall is allowed (Tulving, 1962). But this does not answer the question as to how *nonverbal* items are ordered. However, from everyday observation it is quite obvious that at least for adults there is considerable bias to remember nonverbal items in their original order of presentation. Is this bias acquired by the individual with advancing age or is the young child already predisposed to find the forward order the "natural" one? Administration of a recognition task to child *Ss* is an appropriate method to gain information on this point, since in recognition tasks, unlike recall, there is ordinarily no advantage to be gained by reinstating the original order of item presentation.

When a forward order bias occurs in any type of memory testing, it is very likely that the first item will be reinstated in order more fre-

[1]This research was supported in part by grants from the Social Rehabilitation Services Administration (1484-S) and NICHHD (02026). We would like to thank the school administrators, teachers, and students who kindly cooperated in our work.

quently than any other item, i.e., results will show the traditional primacy effect. Finding such an ordering bias does not advance one far unless variables can be isolated that significantly alter the amount of bias and perhaps in some cases eliminate it altogether. The present two experiments used age and type of material as major independent variables together with a comparison of amount of ordering in spontaneous and required ordering conditions.

The first experiment was conducted within a developmental framework to ascertain how much spontaneous ordering occurs with a minimal number of two items, with three items, and with a choice of two among three items. Specific questions dealt with whether there was an increase with age in a bias toward original order and the extent to which item difficulty affected recognition ordering. With the youngest children, CA 5 it was thought that little or no recognition ordering might occur as children at this age generally lack systematic visual scanning habits such as might be acquired through reading.

The second study was designed based on the results of the first experiment to investigate the amount of voluntary control that children can exert in ordering their recognition responses. In one experimental condition children were *required* to give back their responses in the original forward order; the other experimental condition, used as a control comparison, was like the previous experiment in that the children were free to respond in any order. This comparison allowed determination of the "cost" of ordering; i.e., the amount of loss of recognition accuracy in the required order condition as compared to the free order condition. The cost function is of some theoretical interest as an indicator of the extent to which memory for ordering of items interferes with or is independent of recognition memory for items. In particular, if there is little or no cost in ordering items it can be assumed that order information is relatively independent of item recognition. Cost of ordering was shown clearly in a previous experiment (Postman, Adams, and Bohm, 1956) when two groups of college student *Ss* instructed to learn items in serial order were compared with two groups allowed to learn items in any order. In contrast to the present study, however, individual performance was always far below mastery since their task required retention of a list of 20 nonsense syllables after only four list exposures.

In general, the numerous recent studies on short-term memory have made little attempt to account theoretically for ordering effects such as primacy. The chief current theoretical explanations for forgetting, interference, and decay, each in its own way has considerable to say about why recency effects are found (Brown, 1958; Waugh and Norman, 1965) but tend to ignore primacy. Indeed, Murdock (1968) has shown not only

that primacy effects are widespread, apart from the special case of paired associates, but that a number of recent theoretically oriented experiments deliberately ignore consideration of primacy effects and hence of order-context in general. A developmental determination of the interaction between item ordering and item recognition appears to be a logical first step in investigating memory for order.

EXPERIMENT I

This experiment investigated spontaneous forward ordering for recognition memory by determining the frequency with which an item that was seen first was subsequently recognized first. Three conditions were administered each with an easy type of item, familiar pictures, and a hard type of item, nonsense shapes. In one condition, two of nine items were successively pointed to by *E*, then *S* was asked to recognize them both when the same nine items were again presented (2-2 Cond.). In a second condition, *S* was required to recognize three of nine items (3-3 Cond.). In a third condition, *S* had an option in that three of the nine items were pointed to but only two items were asked for on subsequent recognition testing (2-3 Cond.). It was expected that this latter condition would greatly decrease the bias favoring forward ordering and might eliminate it altogether.

Method

Subjects. The *S*s were taken from an elementary school in Prince Georges County, Maryland, located in a middle-class suburb and a parochial school kindergarten in Washington, D. C. There were 64 *S*s between CA 5-0 and 5-11, 62 *S*s between CA 6-0 and 7-11, and 66 *S*s between CA 8-0 and 9-11.

Materials. There were two sets of nine black and white photos. One set was composed of black on white sketches of nine familiar figures: house, elephant, bed, airplane, umbrella, picnic basket, eye glasses, clock, and necktie with shirt collar. The other set consisted of nine nonsense figures taken from Vanderplas and Garvin's (1959) assortment. Four were four-point shapes: Nos. 1, 9, 13, and 25, while five were six-point shapes; Nos. 3, 4, 7, 15, and 27. These figures were chosen because they were at a midpoint value along the content scale; thus they were neither very high or low in denoting associations with objects or situations. Moreover, they were near the midpoint for their "association value" in the 1959 norms.

Each figure had first been drawn in india ink on a 3×5-inch white index card, then photographed until four clear representations of each of the 18 items were obtained. Each set was mounted in a 3×3 matrix on

hard black printer's board, 15 × 18 inches. There were four mountings of each set, so that within a set each item appeared in a different position. In addition there was a preliminary set of six geometric forms in duplicate. Finally there was an artist's easel, two wooden pointers, and a black cardboard, 15 × 18 inches.

Experimental design. Three conditions were administered: (1) The 2-2 Cond. in which *E* pointed to two familiar pictures (PIX) or nonsense forms (NON) and *S* responded by pointing to two items on the test card; (2) the 3-3 Cond. in which *E* pointed to three items each trial and *S* in turn pointed to three items; and (3) the 2-3 Cond. in which *E* pointed to three items but *S* was asked to respond by pointing to only two items. Each condition was administered to three age groups: CA 5-0 to 5-11; CA 6-0 to 7-11; and CA 8-0 to 9-11. These groups are designated as the CA 5, CA 6-7, and CA 8-9 groups. In the 2-2 Cond. there were 20 *S*s in each of the three groups, there were 24 *S*s in each group in the 3-3 Cond., and 20 *S*s at CA 5, 18 *S*s at CA 6-7, and 22 *S*s at CA 8-9 in the 2-3 Cond. All groups were evenly divided between boys and girls. Half of each sex received a no delay procedure first and half a delay procedure first for each condition.

Procedure. Each child was tested individually. The child stood before the easel, pointer in hand. He was first shown the preliminary or practice sets of geometric forms. The *E* informed the child that he would point to some figures and then the child was to point to these same figures. Practice began with *E* pointing to two of the six displayed figures and the child pointing to these same two figures on the same card. Practice continued until the child successfully pointed to two correct figures arranged in different positions on the duplicate card two successive times. The experiment proper then began.

There were four types of recognition tasks and each child was tested on all four. These tasks required performance with familiar figures and no delay, with nonsense figures and no delay, with familiar figures and a 10-second delay between presentation and recognition, and with nonsense figures and a 10-second delay. There were three trials of each type, hence each child was tested 12 times. Familiar and nonsense figures were alternated each trial. For one-half of the *S*s the first six trials were carried out with no delay and the last six with delay; for the other children the reverse order prevailed.

The procedure on each trial was as follows. The *E* placed two matrix boards on the easel with only the top board showing. He then traced in a rectangular outline either two or three of the nine items on the board as dictated by the particular experimental condition. A metronome set at 2 beats per second was used to control presentation time to about 2

seconds per item. In the no delay condition after the last item was shown, that board was removed and the second board was immediately exposed face up. The child was encouraged to point to the figures just outlined. In the delay condition, after the last item was traced, *E* removed the board but covered the matrix board with a black shield card which remained in place for 10 seconds or 20 metronome beats. At the end of the interval it was removed and the child was requested to respond by pointing to the items that he had previously seen. The *E* recorded all responses on a printed answer sheet.

On each trial the nine items always differed as to position on the two matrix boards. The items were arranged on the response board so that possible scanning habits would be controlled. For example, on six trials no row contained more than one item to be recognized; when three items were to be recognized one row had two and another row one item. The item which was outlined first appeared on the response board an equal number of times in the three rows of the matrix. Thus, if there were a bias to start with the first item shown, *S* would have to begin equally often at the top, the bottom, and the middle of the matrix. Additionally there was no procedural bias favoring left-to-right or top-to-bottom scanning, since 10 of the 12 trials required at least one jump in an opposing direction across rows and columns when *S* pointed to items in the viewed order.

Results

Results in terms of mean percentages of correct recognitions and modal number of items correct are presented in Table 1. Preliminary analyses indicated no effects of delay vs. no delay or sex, so results were collapsed over these variables. With regard to type of material it can be seen that at each age in each condition PIX are better retained than

TABLE 1

MEAN PERCENTAGES OF CORRECT RECOGNITIONS AND MODAL NUMBER CORRECT PER CONDITION ( ) IN EXP. I

CA	Material	Condition		
		2-2	2-3	3-3
5	PIX	88 (2)	82 (2)	67 (2)
	NON	61 (1)	53 (1)	46 (1)
6-7	PIX	87 (2)	82 (2)	78 (3)
	NON	69 (1)	66 (1)	55 (2)
8-9	PIX	97 (2)	92 (2)	91 (3)
	NON	81 (2)	82 (2)	73 (2)

NON; the minimum difference shows a 10% superiority in correct recognition for PIX while the maximum difference is 29%.

The three conditions can also be seen as having produced a consistent outcome. As percentages are compared across rows note that the highest scores are obtained in the 2-2 Cond., the next highest in the 2-3 Cond., and the lowest in the 3-3 Cond. The single exception to this result occurs with NON at CA 8-9 which shows a reversal of 1% between 2-2 and 2-3 Conds.

Assessment of performance as a function of age was made by applying Mann-Whitney U tests between adjacent age groups. Parametric analyses were not justified as the number of responses was small and in some comparisons, Ns were not equal. Reliable increases ($p < .05$) in

TABLE 2

MEAN PERCENTAGES FOR THE CHOICE OF THE FIRST RECOGNIZED ITEM IN THE CONDITIONS OF TABLE 1 AS A FUNCTION OF CONDITION, STIMULUS MATERIAL, AND ORDINAL POSITION

		Condition										
		2-2			2-3				3-3			
		Item position										
CA	Material	1	2	No 1st choice	1	2	3	No 1st choice	1	2	3	No 1st choice
5	PIX	67	19	12	40	17	28	16	36	19	29	15
	NON	41	28	29	23	18	18	42	24	21	17	37
6-7	PIX	70	21	9	37	19	32	11	49	19	22	10
	NON	41	42	18	23	27	19	29	31	28	16	25
8-9	PIX	71	27	3	42	23	24	11	76	11	7	6
	NON	55	35	10	29	35	25	11	42	30	19	10

correct recognitions between CA 6-7 and CA 8-9 were obtained for PIX and for NON material for each of the three conditions. Between CA 5 and CA 6-7 both PIX and NON show significant improvement ($p < .05$) in the 3-3 Cond., only NON in the 2-3 Cond., and neither material in the 2-2 Cond. Note also that improved recognition with increasing age is reflected in the modal number of items correct. At CA 8-9 the modal number is at the ceiling except for NON in the 3-3 Cond.

The extent to which Ss put their recognition responses in the same positional order as they were seen is shown in Table 2. These data refer only to S's first recognition response. This response is analyzed according to whether it was the first, second, or third presented item or whether

it was none of these but instead an item which had not been shown; this last possibility is labeled "no 1st choice."

The primacy tendency is defined by a *S*'s recognizing first the item which had been presented first. In order to evaluate the strength of primacy, intra-*S* comparisons using Wilcoxon signed-ranks tests were made between the first item and either the second, third, or no first choice category, whichever had the highest frequency. For PIX the primacy tendency was predominant at all ages in the 2-2 Cond., at CA 8-9 in the 2-3 Cond., and CA 6-7 and 8-9 in the 3-3 Cond. ($p < .01$, in all cases). With NON material significance ($p < .01$) was obtained only at CA 8-9 in the 2-2 Cond.

Results in Table 2 can be contrasted with those in Table 1 to determine whether the occurrence of significant primacy tendencies is a function of age. Recall that in Table 1 all six comparisons between CA 6-7 and CA 8-9 were significant, and between CA 5 and CA 6-7 three of six comparisons were significant. In Table 2 when Mann-Whitney *U* tests were applied to comparable entries for items in position one, only two comparisons at adjacent ages, both between CA 6-7 and CA 8-9, were significant (NON in the 2-2 Cond. and PIX in the 3-3 Cond., $p < .001$ for both tests). Thus, it appears that overall improvement is more a function of age than is the tendency to favor the first response. This result encourages the interpretation that at younger ages a strong bias is already present for children to order their responses even though they correctly recognize fewer items.

EXPERIMENT II

This experiment compares the spontaneous tendency toward ordering (Free Cond.) with forced ordering when *S* was told that an item had to be recognized in the viewed order to be correct (Required Cond.). The extent to which ordering of items in the Required Cond. exceeds ordering of items in the Free Cond. can be construed as the degree of volitional order control that *S*s can exert. Presumably control of ordering like accuracy of responding should increase with age, but not necessarily to the same extent.

As in the first experiment, different types of stimulus items were administered. In addition to PIX and NON a CF (Color-Form) Cond. was administered in which items could be doubly classified as 1 of 3 colors and 1 of 3 forms. Since this classification system was readily apparent but bore no systematic relation to the items *S* was asked to remember, it can be considered a form of distracting competition rather than an aid. On this supposition it was predicted that CF items would be even harder to recognize than NON items.

Only the 3-3 task was administered and only *Ss* near the extreme ages of Exp. I, CA 6 and 10, were used. As only one type of task was administered, the extent to which *Ss* were able to reinstate order for the second and third as well as the first item could be assessed on a comparable basis for all material.

Method

Subjects. The *Ss* were volunteers who were enrolled at a summer playground in the area of the elementary school attended by the children in Exp. I. Half the *Ss* were at age 6-0 to 6-11 and half at age 10-0 to 10-11. There were four groups of 16 *Ss*, two groups at each age, half boys and half girls.

Materials. The PIX and NON material were the same as those used in Exp. I. Each CF stimulus figure could be classified both by shape and color. The CF material comprised nine discrete figures, three shapes and three colors; shapes were circle, triangle, and crescent; colors were blue, yellow, and red. Figures made of cut-out colored paper were pasted on boards in 3×3 matrices analogous to those used with PIX and NON material.

Procedure. In the Free Cond. the procedure was the same as in Exp. I for the 3-3 Cond. In the required Cond. *Ss* were told that items had to be pointed to in the same order they were seen to be counted as correct. This point was strongly emphasized in the preliminary practice trials. As in Exp. I, half the *Ss* received a no delay and half a 10-second delay procedure. Each *S* was administered a total of nine recognition trials to point to the three items *E* had just pointed to on the previous board. Half the *Ss* in each group were administered three consecutive trials with each type of material in the order PIX, NON, and CF, and half the *Ss* received the reverse order. Positions of items on the response board were arranged to control for possible scanning biases as in Exp. I.

Results

Table 3 reports mean percentages and modal numbers of correct recognitions as a function of age, material, and condition. Sex and delay, which did not yield consistent effects, are collapsed. It must be noted that correct items that were not in order in the Required Cond. are counted as correct in Table 3 even though they would be errors according to the instructions *S* received.

It can be seen that type of material clearly affects correct recognitions. For each of the four independent groups of *Ss* PIX was recognized to a greater degree than NON and NON more than CF material. This rela-

TABLE 3
MEAN PERCENTAGES OF CORRECT RECOGNITIONS AND MODAL NUMBER CORRECT PER CONDITION ( ) IN EXP. II

CA	Condition	Type of item		
		PIX	NON	CF
6	Free	89 (3)	65 (2)	60 (2)
	Required	87 (3)	56 (2)	42 (1)
10	Free	97 (3)	77 (3)	63 (3)
	Required	94 (3)	76 (3)	66 (3)

tive ranking was expected on the basis of the results of Exp. I and from the built-in inter-item competition characterizing the CF material.

In order to assess the cost of ordering items, comparisons were made between the Required vs. Free Conds. Surprisingly, no cost in ordering items was found at CA 10 since Table 3 shows that the means for comparable material are within 3% of each other. Mean recognition scores with both conditions were also almost identical for PIX at CA 6. This similarity of scores did not hold for the other materials at CA 6 as means had differences of 9 and 18% for NON and CF material. Mann-Whitney U tests showed that these differences had probabilities of $<.10$ and $<.03$, respectively. As performance was better with the Free than the Required Cond. with these two materials, some cost was involved in *Ss*' attempt to order responses at this age.

Table 4 presents a more detailed analysis of the recognition means in Table 3 by showing the percentages of correct responses at each ordinal position for the item originally seen in that position. The *Ss* in the Required Cond. had been requested to make recognitions in order while the *Ss* in the Free Cond. had not. Thus, the differences between the Required and Free Cond. at each ordinal position furnish an estimate of the amount of volitional ordering *S* can perform above that done voluntarily without special instructions.

It can be seen in Table 4 that the instructions to order responses in the Required Cond. increased ordering to a greater extent for PIX than for NON or CF materials. With PIX the difference in favor of the Required Cond. attained significance by Mann-Whitney U tests for each ordinal position at both ages ($p < .05$) except for position one at CA 10 ($p < .10$). With NON an expected, though nonsignificant, advantage for the required Cond. is shown for all ordinal positions at CA 10 and all except the third position at CA 6. With CF there is again a uniform advantage for the Required Cond. at CA 10 but no advantage with any

TABLE 4
MEAN PERCENTAGES CORRECT IN ORDER BY ORDINAL POSITION FOR THE CONDITIONS IN TABLE 3

		Type of item								
		PIX item position			NON item position			CF item position		
CA	Condition	1	2	3	1	2	3	1	2	3
6	Required	96	63	58	52	31	23	23	25	23
6	Free	56	29	27	40	29	31	27	29	29
	Difference	40	34	31	12	2	−8	−4	−4	−6
10	Required	88	92	83	65	50	42	56	44	52
10	Free	69	46	38	52	35	35	42	40	35
	Difference	19	46	45	13	15	7	14	4	17

position at CA 6. The negative differences at CA 6 reflect the finding in Table 3 that there is some cost in attempting to order items other than PIX at this age. But, while Table 3 had shown better recognition for the Free than Required Cond., it was unexpected that the children who were explicitly requested to order their responses would, in fact, order items less well than the children who were allowed to recognize items freely.

To test whether primacy ordering was significantly greater than that for other positions, entries for the first position were tested against those for the second or third position, whichever was greater, for the same condition and material. Intra-*S* Wilcoxon signed-ranks tests showed a significant predominance for primacy ordering for PIX at CA 6 in the Required Cond. ($p < .01$) and Free Cond. ($p < .02$), and at CA 10 in the Free Cond. ($p < .02$), but not for the Required Cond. where all responses were highly ordered. Although the first response was the best ordered for all conditions with NON, only the Required Cond. at CA 6 and the Free Cond. at CA 10 showed significant ($p < .05$) primacy ordering. There was a clear absence of significant primacy ordering with CF material for all four groups.

Comparison of the results for ordered responses in Table 4 can be made with the results in Table 3 as to whether either the mean number of correct responses or occurrence of similarity between overall means of the Required compared to the Free Cond. correlates with the incidence of significant primacy ordering. Neither type of result in Table 3 by itself or in combination appears to be a good predictor of Table 4 results.

Finally, it can be noted that cross-comparisons can be made between the 3-3 task with PIX and NON in Exp. I and the Free Cond. in Exp.

II. It is helpful in drawing conclusions to establish reliability where tasks were similar, since in Exp. I a partially different response measure, "choice of the first recognized item," was reported. This measure was appropriate because comparisons of the 3-3 task had to be made with the 2-2 and 2-3 tasks.

In cross-comparisons with regard to mean percentages correct, the Free Cond. entries in Table 3 for PIX and NON at CA 6, 89 and 65, can be compared with the entries 78 and 55 at CA 6-7 in Table 1; likewise, the entries 97 and 77 in Table 3 can be compared with the entries 91 and 73 at CA 8-9 in Table 1. The maximum difference in these comparisons is 11%. In cross-comparisons with regard to mean percentages correct in order, only the entries in Table 2 for the first item position can be meaningfully compared with the entries for the Free Cond. in Table 4. Using the same age cross-comparisons as above, Free Cond. entries for PIX and NON in Table 4 at CA 6, 56 and 40, can be compared with the entries 49 and 31 at CA 6-7 in Table 2; likewise, the entries 69 and 52 at CA 10 in Table 4 can be compared with the entries 76 and 42 at CA 8-9 in Table 2. The maximum difference is 10%. Not shown are percentages correct at item positions two and three in Exp. I. These percentages were calculated and showed a slightly greater discrepancy from Exp. II results with a maximum discrepancy of 17%. As with PIX and NON in Exp. II, the first position was uniformly advantageous with the highest percentage correct at every age including CA 5. It can be concluded that, in spite of inexact age matching and the addition of CF trials in Exp. II, performance was very similar in the two 3-3 tasks.

DISCUSSION

Results from both experiments show that the tendency to reinstate recognition items in their original forward order is strong for children as young as CA 5 or 6 when the primacy tendency is used as a measure of ordering. However, this tendency can be weakened or erased by the introduction of other variables such as the use of nonsense forms as stimuli. Also in Exp. I the 3-3 Cond. proved too difficult to show an ordering effect at CA 5 and the condition allowing optional choices, the 2-3 Cond., showed consistent forward ordering only at the oldest age with the easy PIX material. Results in Exp. II showed that the primacy tendency overestimated the extent of overall ordering in the 3-3 Cond. at younger ages as there was a marked fall-off in ordering of the second and third items in the Free Cond. with both PIX and NON at CA 6.

The NON material was uniformly more difficult than PIX for recognition accuracy and recognizing items in order. In both the Required

and Free Cond. of Exp. II the CF material proved to be the most difficult in terms of overall item recognition and recognition of the first item. This result supports the interpretation that the inherent classificatory organization of the CF material competed with the attempt to order items solely on a chronological basis.

The most interesting comparison was between the Required and Free Cond. of Exp. II. At CA 10 there was no cost in item recognition accuracy in ordering items as both conditions produced a nearly equal number of correct responses. At the same time children in the Required Cond. did, in fact, successfully increase their ordering of items at every ordinal position. The similarity between the Required and Free Cond. could not be a result of the recognition task being too easy since *Ss* made numerous errors.

Comparisons between the Required and Free Cond. at CA 6 were more diverse. With the easy PIX material, results were like those at CA 10. But with NON and CF material the younger children showed some cost in ordering items as more correct recognitions were achieved with the Free than the Required Cond. However, this cost was a result of the *attempt* to order items not of actual item ordering as the younger *Ss* in the Required Cond. did not increase their number of ordered items above the number attained by *Ss* in the Free Cond.

It is proposed that the above comparisons support the idea that item recognition information and item order information are rather independent of each other. They are, of course, not completely independent as order information cannot exist in a vacuum without recognition of individual items. Nevertheless, present results show less than expected interaction between the two types of information. At CA 10 where *Ss* both increased ordering and obtained as many correct recognitions for the Required as the Free Cond., the argument for independence is directly supported. Even the findings at CA 6 with NON and CF material do not imply that the two types of information interfere with each other. If both item and ordinal-position information competed for the same storage space, accuracy of item recognition and ordering should be inversely related, one should decrease as the other increases. This trade-off did not occur. The performance deficit in item recognition at CA 6 appeared to result more from the fact that these younger children did not possess the ability to increase recognition ordering of difficult items on command rather than from a lack of storage capacity.

REFERENCES

BROWN, J. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 12–21.

MURDOCK, B. B., JR. Serial order effects in short-term memory. *Journal of Experimental Psychology,* 1968, **76** (Monogr. Suppl. 4).

POSTMAN, L., ADAMS, P. A., AND BOHM, A. M. Studies in incidental learning: V. Recall for order and associative clustering. *Journal of Experimental Psychology,* 1956, **51**, 334–342.

TULVING, E. Subjective organization in free recall of "unrelated" words. *Psychological Review,* 1962, **69**, 344–354.

VANDERPLAS, J. M., AND GARVIN, E. A. The association value of random shapes. *Journal of Experimental Psychology,* 1959, **57**, 147–154.

WAUGH, N. C., AND NORMAN, D. A. Primary memory. *Psychological Review,* 1965, **72**, 89–104.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 33–39 (1969)

Conjugate Reinforcement of Infant Exploratory Behavior[1]

CAROLYN KENT ROVEE
Trenton State College[2]

AND

DAVID T. ROVEE
Biochemistry Laboratory, Johnson & Johnson Research Foundation

The operant response rate of 10-week-old infants tripled within the first 6 minutes of conjugate reinforcement of foot thrusts, whereas controls receiving identical but noncontingent visual and somesthetic stimulation did not alter response rate throughout the experimental session. All infants receiving the conjugate procedure showed performance sensitive to the reinforcement contingencies for the 27-minute test period; four observations over an extended 46-minute session showed high and stable response rates prior to re-extinction after 42 minutes of continuous testing. The results are interpreted in terms of the feedback characteristics of the procedure.

Over the first 4 months of life, human infants increasingly search and explore their visual environment, the feedback from which is undoubtedly responsible for much of their early and extensive learning (Fantz, 1967; White, 1967). The exploratory motive is well-documented for monkey young and is notable for its strength and persistence in spite of its unspecified biological correlates (Butler, 1953). Piaget (in Flavell, 1963) has defined 1–4 months as the period during which visually guided manual activities emerge. These permit the development during sensorimotor stage 3 of secondary circular ractions, which in turn permit the 4- to 8-month-old infant to alter his environment for its interest value. Similar behaviors have been described by White (1967). Rheingold and her co-workers (Rheingold, Stanley, and Cooley, 1962; Rheingold, Stanley, and Doyle, 1964) found that infants coinciding in age with Piaget's

[1] This research was supported in part by a grant from the Trenton State College Student Executive Board. We thank Lewis P. Lipsitt, Brown University Psychology Department, for his encouragement and critical comments, Carol Ann Kovalich, for assistance with data collection, Arthur Steinman, for programming assistance, and Drs. Charles and Bombet of Baton Rouge, La. for the services of their office in making subjects available.

[2] Department of Psychology, Trenton State College, Trenton, N. J. 08625.

stage 3 are responsive to visual and auditory changes, increasing rate of sphere-touching when it is accompanied by such changes.

Conjugate reinforcement, a variation of free-operant conditioning introduced by Lindsley (1963), has recently been used successfully to demonstrate manipulative learning in year-old children (Lipsitt, Pederson, and DeLucia, 1966). This schedule of reinforcement provides a continuously available reinforcing event, the intensity of which is a direct consequence of response rate. Thus, more rapid responding produces a more intense reward value. Not only does the schedule sustain responding longer and at higher rates than the more typical episodic schedules, but it also permits a continuous record of the efficacy of a reinforcer sensitive to both organismic state changes and learning (Lindsley, 1963; Lipsitt *et al.*, 1966).

In the present study, every attempt was made to provide a reinforcing stimulus which would sustain the attention and behavior of the infants beyond the typical 8- to 10-minute operant test session. The finding of Hunt and Uzgiris (1964) that infants prefer familiar to nonfamiliar mobiles and the evidence that novel stimuli are often preferred to familiar stimuli (Cantor, 1963) suggested that attention might be maximized through a source of "relative" novelty (Berlyne, 1960). This refers to the changing or reorganization of the parts of a familiar stimulus into a "new" stimulus or stimulus pattern. The conjugate reinforcement procedure essentially provided a continuous sequence of "relatively" novel reinforcers, in that the degree or rate of responding produced successive and unique rearrangements of the initially familiar but nonresponsive mobile components.

METHOD

Subjects. Eighteen healthy and apparently normal infants, ranging in age from 9 to 12 weeks, were tested at the time of day reported by the mother to coincide with their normal activity or alert period. The experimental group (four males, two females) had a mean age of 72.5 days; the control *Ss* (seven males, five females) averaged 73.4 days.

Apparatus. All *Ss* were tested in their home cribs with an L-shaped mobile (Nursery Plastics, Inc.) which had been secured to the crib within the first two postnatal weeks. The main features of the mobile were 7 to 10 brightly colored wooden figures suspended on five plastic lines 10- to 12-inches directly above the infant's head. The rigidity of the metal mobile base and the crib weight minimized sway of the figures during normal crib activity. During testing, a soft, silk cord was looped about the left ankle and hooked without slack to the overhead suspension bar.

Procedure. Conjugate reinforcement was provided by means of the ankle cord, in that foot or leg movements directly initiated sway and movement in the mobile figures attached to the suspension bar. Although it was not possible to quantify stimulus parameters, it was apparent that the variety of figure movement increased directly with the force or rate of response. Very rapid responding produced auditory feedback from colliding wooden figures, such that *effectively* more intense responding produced a more intense reward.

Experimental *Ss* received an initial 27-minute session, consisting of a 3-minute baseline period during which the operant level of left foot kicks was established, a 15-minute acquisition phase with conjugate reinforcement, and a 5-minute extinction period. Two minutes intervened between conditions to permit preparation for the succeeding phase. During this interval, *Ss* were permitted to observe the mobile, but leg movement did not effect activity of the mobile, which remained passive throughout.

Control *Ss* received identical conditions for the first and third segments of the procedure; however, during the acquisition segment, all were presented with moving figures continuously activated by the experimenter in a manner simulating the moving figures seen by experimental *Ss* for the 15-minute period. Six of the controls received only noncontingent visual stimulation from the moving figures as described (no ankle cord), and six received noncontingent visual and somesthetic stimulation (cord attached).

A response was defined as a vertical or horizontal excursion of the left foot that returned in the same continuous motion in the direction of origin. Responses were recorded over continuous minutes of testing and out of the direct view of the infant. As a reliability check, a second observer recorded responses independently.

RESULTS

A Spearman rank correlation yielded an interobserver reliability coefficient of .946 ($p < .001$). A lower coefficient obtained over experimental *Ss* only (.82; $p < .01$) probably reflects the difficulty in tabulating all responses when the foot was kicking at the very high rates which characterized conjugate reinforcement periods. This suggests that the response measures may represent conservative records of the reinforcing value of the mobile.

Figure 1 shows mean responses per minute for the experimental group (solid line) and control groups (broken lines). The means of the final 3 minutes of each condition were used in all t tests in order to maximize the opportunity for any effect of the experimental conditions to be

demonstrated. All within-*S* comparisons were indexed by 5 *df*; all between-*S* comparisons, by 10 *df*. The difference between the conjugate reinforcement procedure and the identical but noncontingent visual-somesthetic control group was reliable ($p < .01$), although these groups did not differ significantly ($p > .05$) in baseline or extinction. The attachment of the ankle cord to the mobile did not reliably enhance activity relative to visual stimulation alone ($p > .05$), and neither control group changed from their initial operant levels in successive segments. Within the experimental group *per se*, final acquisition level differed from operant level and extinction ($p < .05$), but extinction did not differ from baseline performance ($p > .05$).

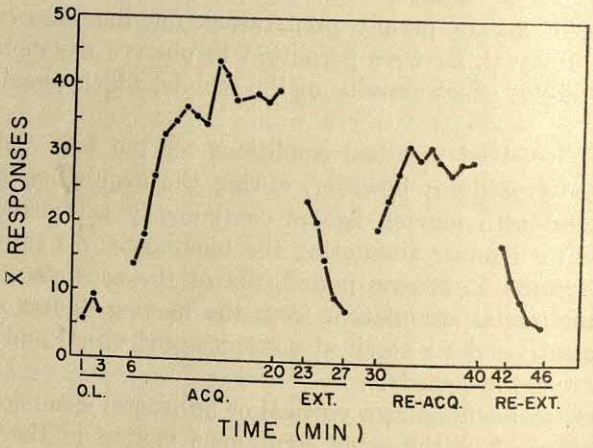


FIG. 1. Rate of response as a function of reinforcement condition. Each *x*-axis point represents a successive minute of observation, approximately 2 minutes separating conditions. The solid line represents the experimental group (conjugate reinforcement); the dotted lines represent control groups (circles = visual-somesthetic controls; squares = visual only). Each curve represents six *S*s.

Four observations were recorded over an additional 17-minute test phase, these infants completing 46 minutes of continuous testing. The additional phase included a 10-minute re-acquisition and a 5-minute re-extinction period, separated by a 2-minute pause (see Figure 2). Comparing mean response levels over the final 3 minutes of each phase as before (see Table 1), performance in each condition significantly differed from that in the preceding conditions with the exception that response levels in extinction did not differ from baseline performance.

Individual data indicated that all infants were clearly under operant control of the reinforcement contingencies. Within 3 minutes infants had doubled the initial response rate, and by 6 minutes of conjugate re-

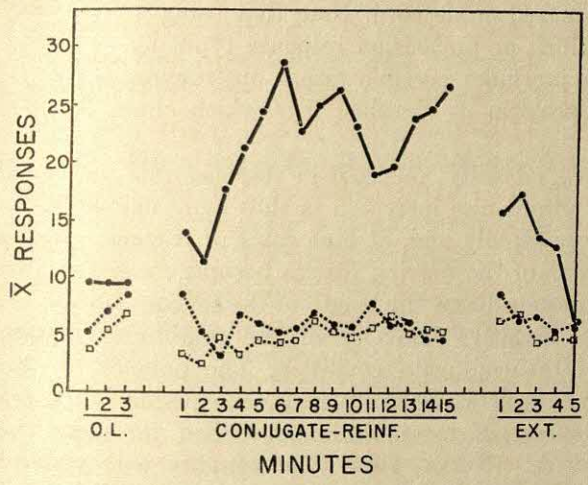


FIG. 2. Mean response rate as a function of reinforcement condition over 46 minutes continuous testing. Note that the reinforcer continued to exert control over behavior even at minute 40, when the response rate was still about 2½ times the operant level ($N = 4$).

inforcement, most had tripled it, both within single test sessions and over repeated sessions. Figure 2 indicates that the effect is clear and reproducible.

TABLE 1
STATISTICAL SUMMARY FOR COMPARISONS BETWEEN FIVE TEST CONDITIONS SHOWN IN FIGURE 2 FOR 46-MINUTE CONTINUOUS TESTING ($N = 4$)

Comparisons	*df*	*t*	Type	Confidence level
Operant level X acquisition	3	4.000	One-tail	$< .025$
Operant level X extinction	3	2.718	Two-tail	$> .05$
Acquisition X extinction	3	3.365	One-tail	$< .025$
Extinction X re-acquisition	3	3.427	One-tail	$< .025$
Re-acquisition X re-extinction	3	2.935	One-tail	$< .05$
Re-acquisition X operant level	3	3.928	One-tail	$< .025$
Re-extinction X operant level	3	0.474	Two-tail	$> .05$

DISCUSSION

The use of the visual-somesthetic control group unequivocally demonstrates that the specific foot thrust represents true learning as a direct result of the contingency between the thrust and the mobile movement and does not reflect merely a state of general arousal produced by the moving target either singly or in combination with somesthetic feedback

from the attached ankle cord. Note that using a foot kick rather than the hand-pulling or prehension response typically described in exploratory studies precludes possible visual obstruction of the reinforcer and evades the problem of clenched fists which characterize this stage of life.

As response intensity, reflected in response rate, increased, the intensity of the reward also increased in that more unique figure movements occurred more rapidly and, at high rates of response, auditory feedback from collisions of the moving figures became more and more prevalent. It may be presumed that the length of the session was achieved primarily through the "relative" novelty of the reinforcing stimulus over the duration of the conjugate condition. The potency of the reinforcing stimulus over such a relatively long time period is reminiscent of the rewarding effects of manipulation described for adult monkeys as a "curiosity drive" (Butler, 1954). The rapidity with which instrumental responses were established (see Figure 1) parallels that reported by Lipsitt *et al.* (1966) for older children and suggests that operant behavior might be effectively conditioned even more in advance of Piaget's stage 3 than has been shown.

The relation between environmental attention and infant learning has been noted in recent reviews (Fantz, 1967; White, 1967). Animal studies and adult human data (Held and Hein, 1963) have suggested that normal development of visually guided spatial perception and coordination depends on the opportunity to observe stimulus variation from self-produced movement. Deprivation studies of institution-reared infants retarded in this aspect of development have shown that simple placement of brightly colored objects in the visual field as early as 37 days of age enhances visual attention (White, 1967). If visual attention is prerequisite to obtaining environmental feedback, hence to learning, then the present study is notable for demonstrating a simple and direct way of achieving and sustaining such attention. Further, the stability of the behavior under the reinforcing contingency, as evidenced after 40 minutes of testing, suggests the usefulness of this technique for the systematic assessment of learning and sensory phenomena in very young infant.

REFERENCES

BERLYNE, D. *Conflict, Arousal, and Curiosity.* p. 38. New York: McGraw-Hill, 1960.

BUTLER, R. A. Curiosity in monkeys. *Scientific American Offprint,* 1954 (Feb.), No. 426.

BUTLER, R. A. Discrimination learning by rhesus monkeys to visual exploration motivation. *Journal of Comparative and Physiological Psychology,* 1953, 46, 95–98.

CANTOR, G. N. Responses of infants and children to complex and novel stimulation. In L. P. Lipsitt and C. C. Spiker (Eds.), *Advances in Child Development and Behavior*, Vol. I, pp. 1–30. New York: Academic Press, 1963.

FANTZ, R. L. Visual perception and experience in early infancy: A look at the hidden side of behavior development. In H. W. Stevenson, E. H. Hess, and H. L. Rheingold (Eds.), *Early Behavior*. Pp. 181–224. New York: Wiley, 1967.

FLAVELL, J. *The Developmental Psychology of Jean Piaget*. Pp. 85–109. Princeton, N. J.: Van Nostrand, 1963.

HELD, R. M., AND HEIN, A. V. Movement-produced stimulation in the development of visually-guided behavior. *Journal of Comparative and Physiological Psychology*, 1963, 56, 872–876.

HUNT, J. MCV., AND UZGIRIS, I. C. Cathexis from recognitive familiarity: An exploratory study. Symposium paper presented at meetings of the American Psychological Association, Los Angeles, California, 1964. Pp. 1–15.

LINDSLEY, O. R. Experimental analysis of social reinforcement: Terms and methods. *American Journal of Orthopsychiatry*, 1963, 33, 624–633.

LIPSITT, L. P., PEDERSON, L. J., AND DELUCIA, C. A. Conjugate reinforcement of operant responding in infants. *Psychonomic Science*, 1966, 4, 67–68.

RHEINGOLD, H. L., STANLEY, W. C., AND COOLEY, J. Method for studying exploratory behavior in infants. *Science*, 1962, 136, 1054–1055.

RHEINGOLD, H. L., STANLEY, W. C., AND DOYLE, G. Visual and auditory reinforcement of a manipulatory response in the young child. *Journal of Experimental Child Psychology*, 1964, 1, 316–326.

WHITE, B. L. An experimental approach to the effects of experience on early human behavior. In J. P. Hill (Ed.), *Minnesota Symposia on Child Psychology*, Vol. I, pp. 201–226. Minneapolis: University of Minnesota Press, 1967.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 40–44 (1969)

Computer Analysis and Display of Movement Patterns[1]

R. E. HERRON[2] AND M. J. FROBISH[3]
University of Illinois

A description is given of an automatic method for producing numerical and graphical statements about movement and inter-organism or organism-environment interaction patterns from observational records. The method is based on use of Cartesian co-ordinates to represent the position of each individual at each instant sampled. In addition to statistical computations regarding distances travelled, the frequency of entry into specified territories, sociometric and other parameters, movement maps can be produced with the aid of a plotting device connected to a computer. The program can be adapted for various investigations concerning social interaction, territoriality, equipment usage, activity habits, and architectural design.

In 1960, Wright warned that progress in realizing the potential of electronic observational child study would "depend on development of analysis procedures to suit the new refinement and fidelity of its records [p. 124]." Today, this need is vividly apparent and, as access to simply operated electronic observational aids continues to grow, so will the backlog of unanalyzed visual records of child behavior continue to build up until suitably refined, automatic methods of analysis are widely available.

A special problem exists with studies involving activity observations over periods longer than a few minutes because they tend to produce large quantities of position data which ultimately, have to be reduced to meaningful statements about the *patterns* of movement or interaction

[1]This research was supported in part by grant MH-07346 from the National Institute of Mental Health and a grant (Project 9424) to the first author from the Illinois State Department of Mental Health. The authors thank R. J. Korb who was responsible for selecting the photographic equipment and processing the photographs; L. Wuellner, Research Assistant, who carried out the observations; M. J. Ellis, Research Associate, and P. A. Witt, Research Assistant, for their helpful suggestions, and J. Kolman, Statistical Consultant, who contributed advice on programming. They are also indebted to the staff, parents, and children (the subjects) of the Children's Research Center Preschool at the University of Illinois for their kind cooperation, and Drs. W. P. Hurder and A. V. Sapora who helped in various ways to make this research possible.

[2]Now at Baylor University College of Medicine, Texas Medical Center, Houston, Texas 77025.

[3]Now at Bradley University, Peoria, Illinois 61606.

involved before they can be interpreted. When the analysis has to be carried out manually, as often is the case, it is a tedious and time-consuming task—a state of affairs which does nothing to encourage an exhaustive analysis.

The method described below grew out of efforts to automate a series of studies of children's movement patterns, social interaction, territoriality, and equipment usage in an indoor play area. The purpose of this report is to outline the principle upon which the method is based and to discuss some of the potentialities and limitations.

METHOD

Data Collection

Spatio-temporal position data were obtained by means of a time-lapse photographic system; this system was chosen in preference to film or videotape, or electronic devices such as ultrasonic or radio transmission, on account of its relatively low cost, ease of maintenance, and simplicity of operation. A remotely controlled *Nikon* F camera was semipermanently mounted in the ceiling, 9 feet above a 29½- × 22-foot indoor play area in the Children's Research Center of the University of Illinois. The camera was equipped with an accessory *Accura* fish-eye lens fitted to a 50-mm, *f* 1.4, Nikkor lens and an accessory 250-exposure motor drive so that an extended sequence of time-lapse exposures could be taken automatically. The 7.5-mm focal length of the combined lenses takes in a true hemisphere and the entire play area was within the field of view. In a recent experiment, the photographs were taken at intervals of 10 seconds throughout a series of 15-minute play sessions, but the system includes an intervalometer which provides for varying the time between exposures over a wide range.

The only manual step in the data acquisition process was the identification of each child's position on each photograph in terms of Cartesian co-ordinates. A grid of three-foot squares which shows up distinctly in the photographs was used to obviate the need for manually superimposing a co-ordinate scale over each photograph. The child's position was taken as the point on the floor which was estimated to be directly below the child's center of gravity. This step was carried out manually for want of a suitable and inexpensive automatic procedure—the semi-automatic method previously described by Haith (1966) in this journal is not too feasible when a fish-eye lens is used for taking the photographs. Areas or "territories" enclosing certain pieces of fixed equipment were also identified by Cartesian co-ordinates. These data when transferred to IBM cards served as the input to the computer-plotter.

Computer Analysis and Plotter Display

Once the position data are in the form of Cartesian co-ordinates the computer-plotter can be programed to yield various permutations of the following numerical and graphical data:

1. Each child's changes in position during each play session displayed graphically, in chronological order, on a plan view of the play area (Fig. 1).
2. The distance represented by the total changes in position of each child per play session and the mean distance for each child over all sessions.
3. The aggregate distance between each child and each of the other children for each exposure and for each play session.
4. The mean distance between each child and each of the other children during each play session computed and displayed as a function of time (i.e., with succeeding play sessions).
5. The frequency of entry of each child into each of the specified territories for each play session computed and displayed as a function of time.
6. The frequency each child is alone in each territory per play session and the incidence of solitary activity computed and displayed as a function of time.
7. The frequency each child is accompanied by selected permutations of other children in each territory per play session computed and displayed as a function of time.
8. The aggregate frequency of entries by all children into each of the specified territories for each play session computed and displayed as a function of time.

All of the above output components have yet to be incorporated into a single program and this may seldom be necessary because of the considerable breadth of inquiry which such a program would represent. Suffice it to say that these computations and displays are within the capacity of current computer-plotter combinations and that additional parameters or more elaborate statistical treatments can be added with little difficulty.

The graphical displays shown in Figure 1 were produced with the aid of a program written in *Fortran* II for an *IBM* 7094 computer and a *CalComp* 670-564 (*California Computer Products*) plotter. The figure is composed of two "movement maps," drawn automatically by the plotter, which indicate the positions and changes of position of two children during the same play session, with Child A's map on the left and Child

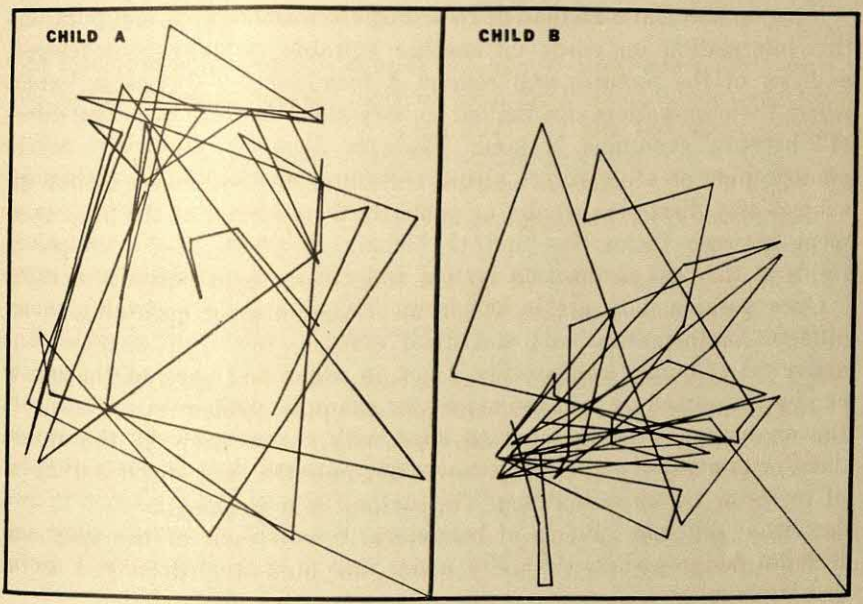


FIG. 1. Corresponding movement patterns of two children drawn automatically by a computer-plotter. (For explanation, see text.)

B's map on the right. Each map covers the same indoor play area. The area was liberally distributed with various pieces of play equipment, such as large hollow blocks, ladders, and a see-saw. The apparatus was omitted here in order to simplify reading of the map, but the relative positions of immovable objects can be added if desired, by means of a template. Differences between the two children in their focal points of interest and their movement patterns during the period in question show up quite distinctly.

DISCUSSION

The above method has certain fairly obvious limitations—an important one is illustrated in the above examples. Movements between the points (showing the child's position every 10 seconds) which the plotter joins up with straight lines to form the map are completely missed. This illustrates the need to carefully consider such matters as the sampling interval and the use of random versus regular intervals. Such decisions will be influenced greatly by the exigencies of each experimental situation as will the reliability and validity of the method. But, preliminary findings indicate that, with reasonable care, high levels of reliability and validity are attainable.

Until an automatic method of recording co-ordinate data and punching this information on cards or another suitable medium is developed, analysis of the pictures will remain a formidable undertaking except where the observations are limited to very short periods or a long interval between exposures is used. There is also the limitation, where photographic or other direct visual recording media are used, that the subject may disappear wholly or partially from view (by getting equipment between themselves and the lens). However, these limitations reside in the data acquisition system and not in the analysis procedure.

Once position data are in the form of Cartesian co-ordinates, many different mathematical and statistical manipulations are possible. The above list of output components is not an exhaustive one; new elements can be programed as the need arises, for example, with some adjustments the analysis can be adapted to cope with three-dimensional position data, or to analyze and display movement patterns throughout a complex of rooms or an entire building. The method is now being applied in two somewhat different spheres of behavioral research—children's play and hospital design—where the users' claim that it is proving to be a useful research tool.

REFERENCES

HAITH, M. M. A semi-automatic procedure for measuring changes in position. *Journal of Experimental Child Psychology*, 1966, 3, 289–295.

WRIGHT, H. F. Observational child study. In P. H. Mussen (Ed.), *Handbook for research methods in child development*. New York: Wiley, 1960. Pp. 71–139.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 45–62 (1969)

Increasing Task Behavior in a Language Arts Program by Providing Reinforcement[1]

DONALD E. P. SMITH, DALE BRETHOWER, AND RAYMOND CABOT
Center for Research on Language and Language Behavior, University of Michigan

The work of retarded readers using a programmed language arts curriculum was observed under various conditions of reinforcement in a controlled classroom. Task behavior was followed by conditions of no consequence, teacher praise, a work-break consequence, a monetary consequence and feedback on amount of work (relative to previous work). Extinction of task behavior tended to occur under conditions of no consequence and of teacher praise. Conditions of monetary consequence and of feedback on progress resulted in high, sustained rates of work behavior.

Maintenance of task behavior is a chronic problem in classrooms and clinics. Common pedagogic terms for classifying the problem are "inattention," "lack of motivation," and "undisciplined behavior."

Common classroom techniques for increasing task behavior include teacher praise, withholding punishment, assigning grades and inducing a commitment by means of a contract. This paper reports the results of several experiments designed to increase task behavior in a controlled environment, i.e., in classrooms wherein instructional materials and teacher behaviors are controlled and wherein the dependent variable is rate of work, i.e., correct task responses per hour.

The problem of maintaining task behavior arose during the validation of a programmed language arts curriculum.[2] The program, designed for use in first grade classrooms and in clinics, consists of 16 books containing some 17,000 tasks, scripts (or tapes) for teacher participation, and a programmed manual for training the teacher in his role. The program trains reading, writing, listening, and speaking behaviors with

[1] The research reported herein was performed in part under Contract OEC-3-6-061784-0508 with the U. S. Department of Health, Education, and Welfare, Office of Education, under the provisions of P.L. 83-531, Cooperative Research, and the provisions of Title VI, P.L. 85-864, as amended. This research report is one of several which have been submitted to the Office of Education as *Studies in language and language behavior,* Progress Report V, September 1, 1967.

[2] Smith, D. E. P. (Ed.) *Michigan Language Program.* Ann Arbor; Ann Arbor Publishers, 1964, 1965, 1966, 1967.

tasks ranging from matching letters and utterances, initially, to reading paragraphs orally and silently as terminal behaviors. The learning tasks were constructed iteratively, within classrooms, until an error rate below 5% was achieved on each book.

The teacher's role is that of classroom manager. He arranges the classroom, provides tasks and materials, and provides feedback on performance. Performance feedback is carefully controlled and predictable. Under normal operating conditions, feedback on task relevant performance is provided by the programmed materials which are self-selected and self-paced. Feedback on non-task performance is provided by establishing a limited number of rules—usually one. Non-task performance not covered by the rule is followed by no consequence other than being noticed; performance covered by the rule is always followed by a specific enforcement procedure, i.e., stating the child's name and requesting a restatement of the rule ("John, what is the rule?").

Within these limits, children are free to work or not to work. It should be noted that most of the teaching frames are of the discrimination type and control the attention of most children in a stable, predictable environment. A "frame" in this program is defined as a task requiring a discriminative response or a production. A response, then, is defined as completion of a single task—circling a letter, indicating same or different, spelling a word, writing a dictated sentence, etc. Rate of work is determined by counting the graphic evidence of responses, either the trained response (such as the spelling of a word) or a "token" response (such as circling).

Rate of work in normal classrooms under the above conditions tends to be high (i.e., 200 correct responses per hour). However, teacher reports of extinction curves for isolated cases were confirmed by similar data on a substantial proportion of clinic cases.

In an attempt to resolve the problem of maintaining task behavior, a series of six experiments was carried out. The independent variable in each was none, one or a combination of consequences following task behavior. The first two experiments demonstrate the results of no consequence and of a monetary consequence precisely contingent upon amount of task behavior. The next three experiments report results under conditions approximating each of these extremes. The last study reports the effect of a non-monetary consequence feasible for school use.

Experiment 1: No Consequence

Procedure. Ss were six children designated "non-verbal, first-grade failures" by their school principal (Table 1). Ages ranged from 70 months to 98 months. Reading skill deficiencies ranged from 1 month to

6 months (Gates *Primary:* Word Recognition). Classes met 50 minutes per day, 4 days per week for 10 weeks. Conditions and materials were as described above. No consequence followed work behavior.

Results. The cumulative curves depicting rate of work (Figure 1) characteristically demonstrate extinction of task behavior. Reduced outputs were usually accompanied by emotional behaviors, attacking self and others, "escape" responses, curling in a fetal position, sitting under the desk, etc. Under these conditions, mean rate of work was 39 responses/hour.

Discussion. The condition of no consequence following task behavior results in a low rate of work, rapid extinction of task behavior and a substantial amount of infantile attack and withdrawal behavior.

TABLE 1
WORK OUTPUT OF RETARDED READERS UNDER CONDITIONS OF NO CONSEQUENCE FOR TASK RELEVANT BEHAVIOR

Subject	Sex	Age (months)	Grade placement	No. of sessions	Responses	
					Total	$\bar{x}$
A	M	98	1.9	21	1650	78
B	M	70	1.9	26	1100	42
C	M	88	1.9	24	800	33
D	F	71	1.9	7	500	71
E	M	70	1.9	25	600	24
F	F	66	1.9	27	350	13

Experiment 2: Monetary Consequence, Task Contingent

Procedure. *S*s were six children taken in order from a waiting list and assigned to a class taught by the teacher used in Experiment 1. Ages ranged from 101 months to 136 months. Reading skill deficiencies ranged from 1.1 years to 2.7 years. Children were described by their teachers as hyperactive (2), anxious (1), passive-aggressive (2), aggressive, acting-out (1) (school placement in a classroom for disturbed).

Classes were held for 45 minutes, 4 days per week, with the same physical arrangements and materials as before. Two rules obtained, "No talking during independent work," enforced as above, and "No disturbing others," enforced by exclusion for the remainder of the hour. Immediately following the class, correct responses were totaled, in the presence of the child, and a monetary pay-off occurred. The child was paid at a rate of one cent for ten responses by the office secretary.

Results. Rate of work began at 100 responses per session and reached a mean of 430 responses per session during the last four meetings

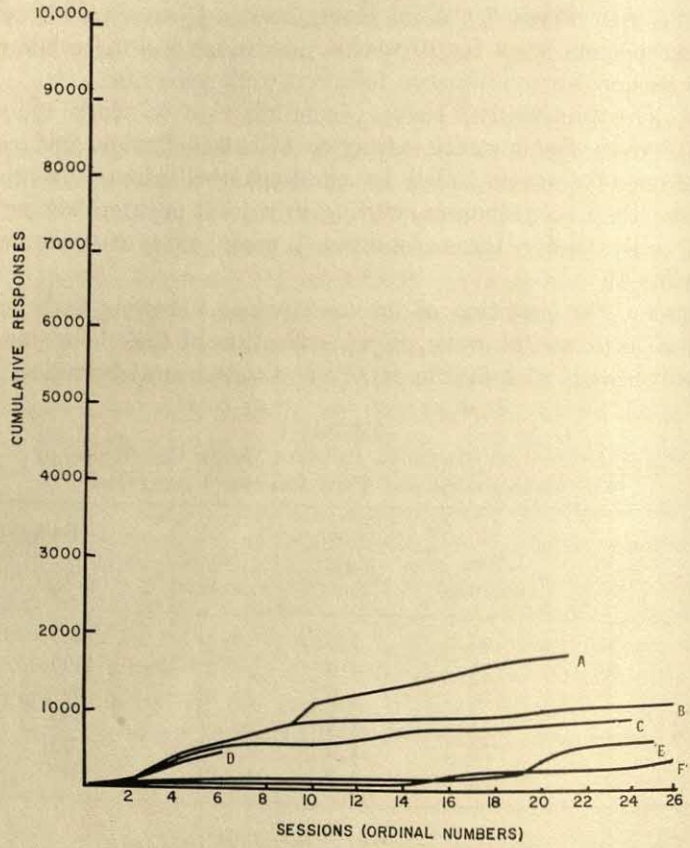


FIG. 1. Work output of retarded readers under conditions of self-selection, self-pacing and no consequence.

TABLE 2

WORK OUTPUT OF RETARDED READERS UNDER CONDITIONS OF A MONETARY CONSEQUENCE CONTINGENT UPON NUMBER OF TASKS COMPLETED CORRECTLY

Subject	Sex	Age (months)	Grade placement	No. of sessions	Responses	
					Total	$\bar{x}$
A	M	117	4.6	24	7000	292
B	M	124	4.6	24	6400	267
C	M	136	5.6	20	5775	289
D	M	112	4.6	23	4600	200
E	F	101	3.6	18	3825	212
F	M	114	3.6	16	3400	213

(Figure 2). All curves are accelerating positively by Session 16. Mean proportion of time spent in task behavior exceeded 90% on three of the last five sessions.

Gains in reading skill are reported in Figure 2A. Test data indicate no change over the 6 weeks during which rate of work data was recorded. Five of the six *Ss* continued treatment through 14 weeks by which time they had completed the program. Whereas mean deficiency was two grades and remained so after 6 weeks, mean final deficiency is slightly less than .6 grades or an achievement level of grade 4.0, the upper limit of the program as now constituted. "Expected achievement" is based upon mean chronological age. The mean gain of 1.4 grades may be compared with a median gain of .4 grades achieved by a sample of 50 similar children treated in the year prior to this study.

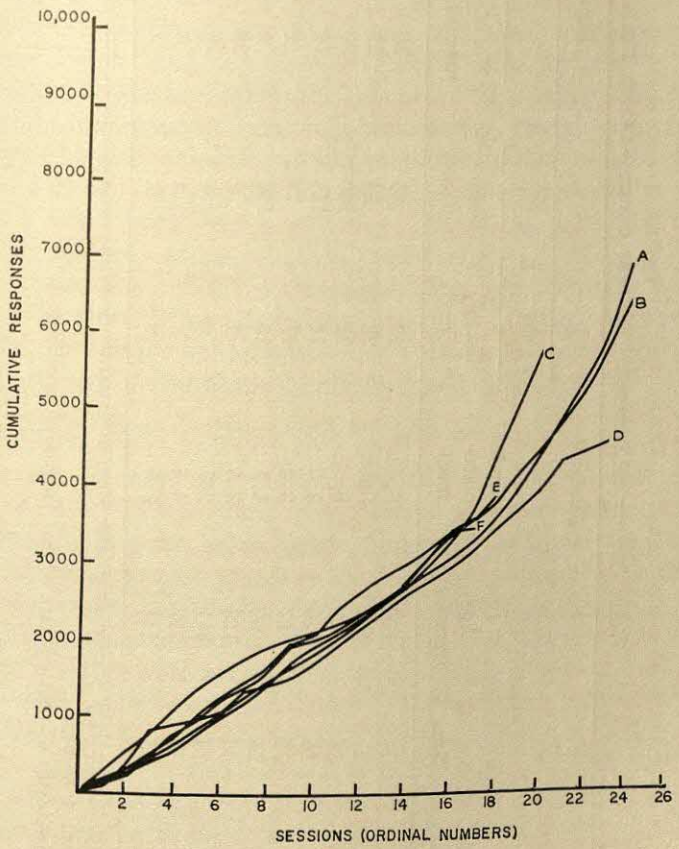


FIG. 2. Work output of retarded readers under conditions of self-selection, self-pacing, and a monetary consequence.

TABLE 3

WORK OUTPUT OF RETARDED READERS UNDER VARIOUS TREATMENT CONDITIONS

Subject	Sex	Age (months)	Grade placement	Experiment[a] 3 No. of sessions[b]	3 Mean responses	4 No. of sessions	4 Mean responses	5 No. of sessions	5 Mean responses
A	F	106	Spec.	23	36	20	19	23	26
B	F	102	Spec.	23	74	23	32	24	83
C	M	120	Spec.	21	160	20	100	8	188
D	M	110	3.0	14	118	19	101	15	152
E	M	97	3.0	17	75	x	x	x	x
F	M	99	4.7	16	119	x	x	x	x
G	M	103	3.5	x	x	20	119	21	230
H	M	111	2.7	x	x	16	169	20	348
I	M	127	4.5	x	x	x	x	23	246
J	M	141	5.5	x	x	x	x	23	157

[a] Exp. 3: Contract + teacher praise; Exp. 4: Contract + no consequence, then work-breaks as consequence, then money; Exp. 5: Contract + money.

[b] x Indicates no participation in that treatment condition.

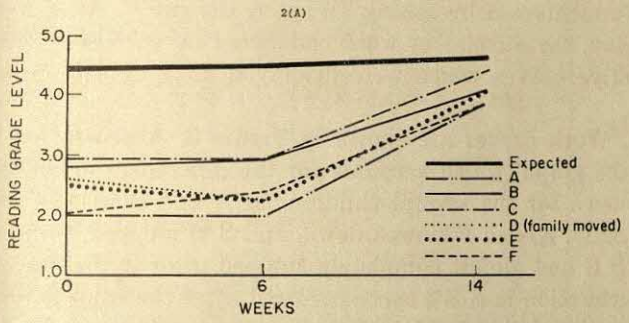


FIG. 2A. Reading achievement of clinic referrals after two periods of treatment with programmed materials in a controlled environment.

Discussion. The condition of a monetary consequence contingent upon task behavior was clearly effective in producing a high rate of work and systematic achievement. There was virtually no rule-testing after ten sessions with this group of relatively disturbed children.

Subjective evidence during the course of the experiment suggested that, while the apparent reinforcer was money, the effective reinforcer might well be feedback information on progress. For example, one child received $.26 one day and $.05 the next. He said, wonderingly, "H-m-m. I guess I didn't work very hard today." It was as though he had been previously unaware of his role as a producer.

If feedback is an effective reinforcer, other non-monetary procedures might be as powerful as the money pay-off. A non-monetary procedure might be feasible for public school use. Experiment 6, described later in this report, uses performance feedback rather than money.

Experiment 3: Work Contract with Teacher Praise

Procedure. Ss were six children referred by their school principals (Table 3, A–F). Three were from special education classes, two read approximately 1 year below grade placement and one read at grade level. The boy who read at grade level was not performing at that level in school and was described as a behavior problem.

Conditions were similar to those described for Experiments 1 and 2. However, children were asked to complete small amounts of work. Agreement to work constituted a "contract." When work contracts were completed, the child brought his work to the teacher who smiled and made comments such as, "Good! You finished. Would you like to do some more?" If the child agreed to do more, the next contract was set. If the child did not agree to do more, he was instructed to go to a non-work area containing play materials.

Rules were enforced by asking "What is the rule?" After the children left each day, the amount of work completed by each child was plotted on cumulative curves which were displayed on a bulletin board in the classroom.

Results. Work curves are shown in Figure 3. Absences are not indicated on the graph, which accounts for the differences in the number of sessions shown for the several children. *Ss* A and B worked slowly and steadily; *Ss* D, E, and F slowed down, and *S* D stopped, then went back to work; *S* E had almost completely stopped prior to Session 10.

The acceleration in *S* E's curve occurred after the student was allowed to skip exercises if he could read the summaries at the end of exercises. After completing the exercises where that was possible, he stopped working, whereupon he was tested (Session 18) and graduated. He was then

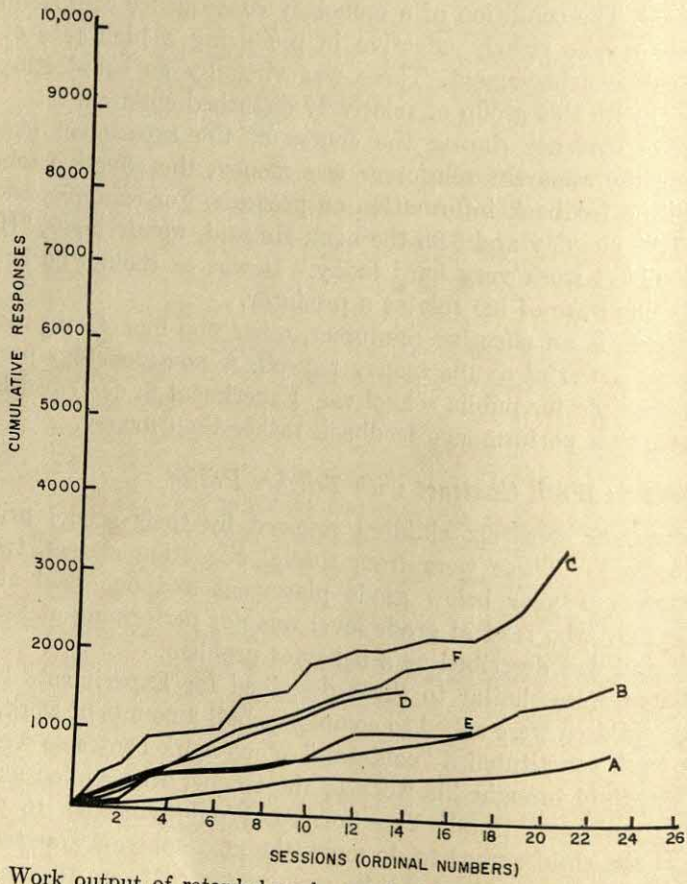


FIG. 3. Work output of retarded readers under conditions of work contracts with teacher praise.

reading above grade level. Until Session 18, curve C has the general appearance of several small extinction curves superimposed upon a larger extinction curve. The relatively large number of responses in Session 7 occurred as the child stayed beyond the class session to complete a rather large program book. A similar event occurred in Session 13. The acceleration following Session 18 occurred when Ss D and F were absent often and E had graduated, reducing the effective class size to three students. Most of S C's work during these sessions was on teacher directed and teacher paced exercises.

Discussion. With one exception, the consequence of teacher praise resulted in rates of work slightly higher than those produced under Experiment 1, No Consequence, but work behavior was not well maintained. Much of the time the children were engaged in activities such as running (quietly) around the room, tapping pencils, etc. Although considerable work was accomplished, working was clearly not the major activity in the classroom. The work curves displayed in the classroom were rarely examined by the children.

Experiment 4: Work Contract with No Consequence, Work-break Consequence and with Monetary Consequence

Procedure. Ss were six children; four of them, A, B, C, and D, had taken part in Experiment 3; two of them (G and H) were new, replacing the child who had graduated and one child who was not re-enrolled for the 6-week period (Table 3). Classes met 1 hour per day, 4 days per week for 6 weeks in the same classroom used in Experiment 3.

No rules were used; the teacher allowed movement, talking, etc., and requested halts or reductions in non-task activity whenever he judged it necessary. Upon completion of a work contract or a work break, the child received the option of another contract or a work-break. Beginning with the 10th session (see (1) in Figure 4), the option was between another contract or a maximum break of 5 minutes; two consecutive breaks were not allowed. Beginning with Session 16 (see (2) in Figure 4), pennies were given to the children upon completion of each contract. The amount per correct response was not specified to the child. He might receive one penny per response upon completion of one contract, one penny per two responses on the next. However, once the ratio of correct responses to pennies was increased, it was not decreased. The maximum ratio was one penny to four responses. Cumulative curves of each child's performance were plotted after class and displayed in the classroom.

Results. Work curves are shown in Figure 4. (The breaks in curves which occur just before procedure changes are due to absences which may have occurred any time the particular procedure was in effect;

however, curves are aligned at the beginning of each condition in order to facilitate representation of procedure changes.)

There is no general decline in work (and no teacher-paced acceleration). However, prior to the introduction of a "work-break" consequence, there is a deceleration of the four upper curves.

Following the introduction of "work-break" consequences, curves C, D, G, and H accelerated and then decelerated. Curves A and B show no change; the children rarely came in contact with the "work-break" consequences, partly because of low frequency of completing contracts and partly because they occasionally exceeded the time allotted for the break.

Five of the curves show acceleration after the introduction of monetary consequences. The acceleration in curve C is quite noticeable.

Discussion. Ss A and B did little with the money except to ask what

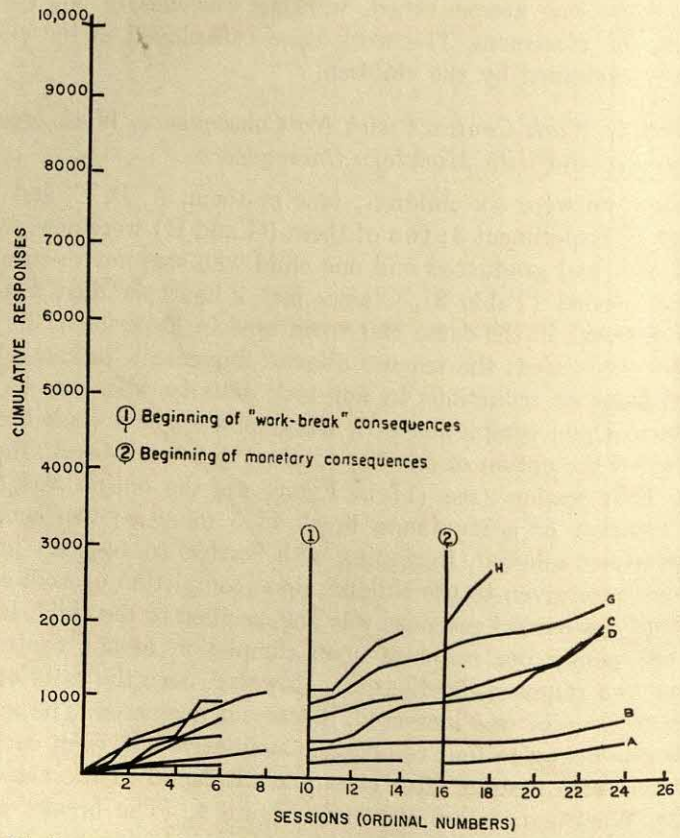


FIG. 4. Work output of retarded readers under conditions of work contract with no consequence, with work-break consequence, and with monetary consequence.

it was for. After Session 17 the teacher suggested to the parents of the two children that they encourage the children to spend some of the money immediately after class. The subsequent acceleration in the curve for child B may have been due to a resulting increase in the value of the money to the child.

Experiment 5: Work Contract with Monetary Consequence Plus Progress Plotting

Procedure. Procedures were similar to those used in Experiment 4. Ss were eight children A, B, C, D, G, H, I, J, described more fully in Table 3. Child B was in a different class operated under the conditions described for Experiment 2. Children I and J joined the class after having been a two-person class for a 6-week period. Child C was scheduled to attend the class only on Tuesday and Thursday.

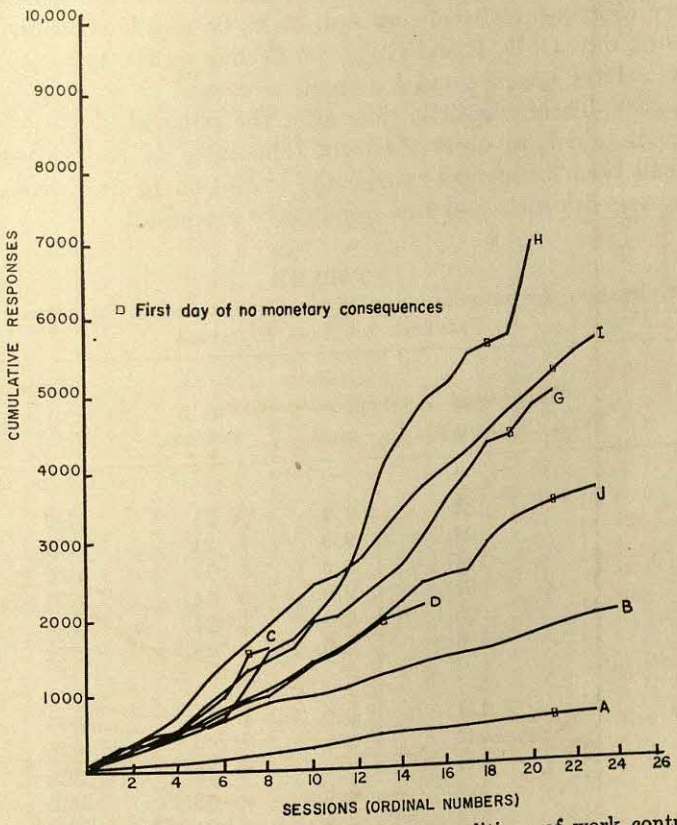


FIG. 5. Work output of retarded readers under conditions of work contract with monetary consequence and work contract with no monetary consequence.

The class was taught part of the time by a male teacher and part of the time by a female teacher.

The work-to-payment ratio was not stated to the children. Beginning with Session 11, the ratio was systematically increased. In Session 11 it was 5/1, in Sessions 12–15 it was 10/1, in Sessions 16–19 it was 15/1, in Session 20 it was 20/1, in Session 21 it was 25/1. There was no monetary consequence during Sessions 22–24. Students graphed their work output at the end of each session throughout the experiment.

Results. The cumulative work curves are shown in Figure 5. Absences are not indicated on the graphs. The performance was maintained throughout the experiment. There is some deceleration in curves J and H as the ratio of correct responses per penny was increased from 5/1 to 25/1 but the other curves are maintained or even accelerate at the higher ratios.

Curves A, D, I, and J show no noticeable effect on the first day monetary consequences were removed. *Ss* H, G, and C show an increase on the first day. C, D, I, and G show a decline on the last day whereas H shows a large increase and J a small increase.

Discussion. There was little time after the removal of monetary consequence in which to observe effects. The ratio of correct responses/penny had been increased to make the transition to no monetary consequence less dramatic and was apparently successful.

TABLE 4

AGE, READING ACHIEVEMENT AND WORK OUTPUT UNDER CONDITIONS OF FEEDBACK ON TASK BEHAVIOR

Subject	Sex	Age (months)	Grade placement	No. of sessions	Responses	
					Total	$\bar{x}$
I						
A	M	132	4.9	21	5740	274
B	M	98	2.9	21	5725	273
C	M	124	3.9	19	5150	271
D	F	97	2.9	15	2775	185
E	M	98	2.9	21	3800	181
F	M	106	2.9	18	2750	153
II						
A	M	151	6.9	17	8175	481
B	M	144	5.9	20	9875	494
C	M	141	6.9	13	4925	379
D	M	128	4.9	18	5425	301
E	M	122	4.9	21	4700	224
F	M	140	5.9	16	2650	166
G	M	109	3.9	21	3475	165

The average response rates shown in Table 3 show a consistent pattern: In all cases where the comparisons can be made, a child's response rate was lower in Experiment 4 than in Experiment 3 and higher again in Experiment 5. Our interpretation of this pattern is that the first 16 sessions of Experiment 4 represent a continuation of the decline in rate observed in Experiment 3. That is, teacher praise and work-break consequences were not sufficient to maintain the responding of the children who took part in both experiments.

Experiment 6: Non-Monetary Consequence, Task Contingent

As outlined above (Experiment 2: Discussion), if the consequence, feedback of information, is an effective reinforcer in Experiment 2, other, non-monetary kinds of information may maintain task behavior. The simplest of these kinds of feedback appears to be a report of the

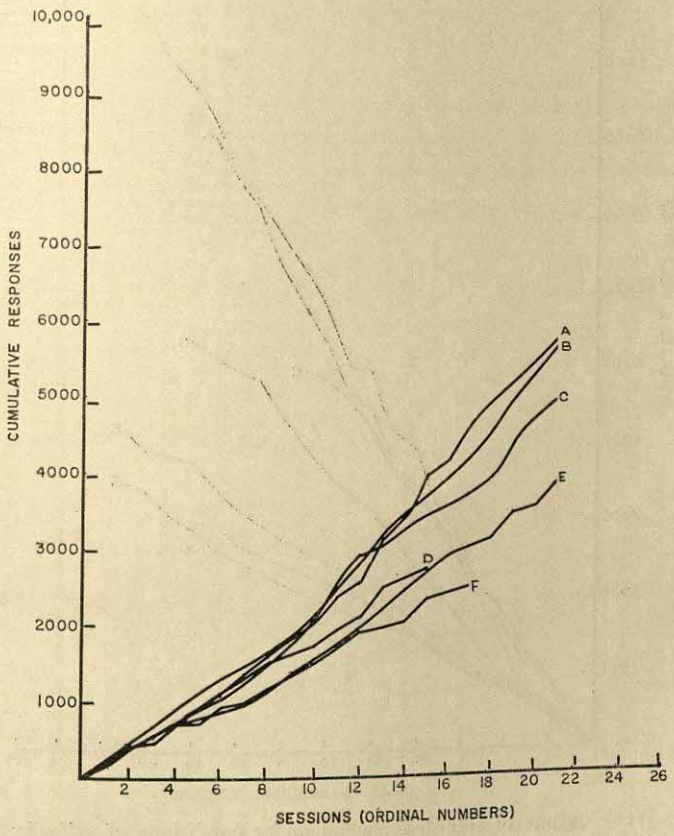


FIG. 6. Work output of retarded readers under conditions of self-selection, self-pacing, and immediate feedback on performance quantity (Class I).

number of tasks completed, perhaps charted in a way which will allow a daily and weekly comparison of output.

Procedure. *Ss* were 13 children referred by their principals for treatment. Six were placed in Group 1 with a teacher experienced in operating a controlled environment. Ages ranged from 8–1 to 11–0. Reading deficiencies ranged from .0 grades to 2.3 grades (Table 4). One boy, F, was classified as retarded with special room placement. B and E are twins.

Seven boys were placed in Group II with a teacher recently trained in operating the room. Ages ranged from 9–1 to 12–7. Reading deficiencies ranged from 1.9 grades to 4.5 grades. Two boys were classified as retarded with special room placement, F (WISC IQ, 82) and G (WISC IQ, 69).

Materials and procedures were those of Experiment 2. The two rules

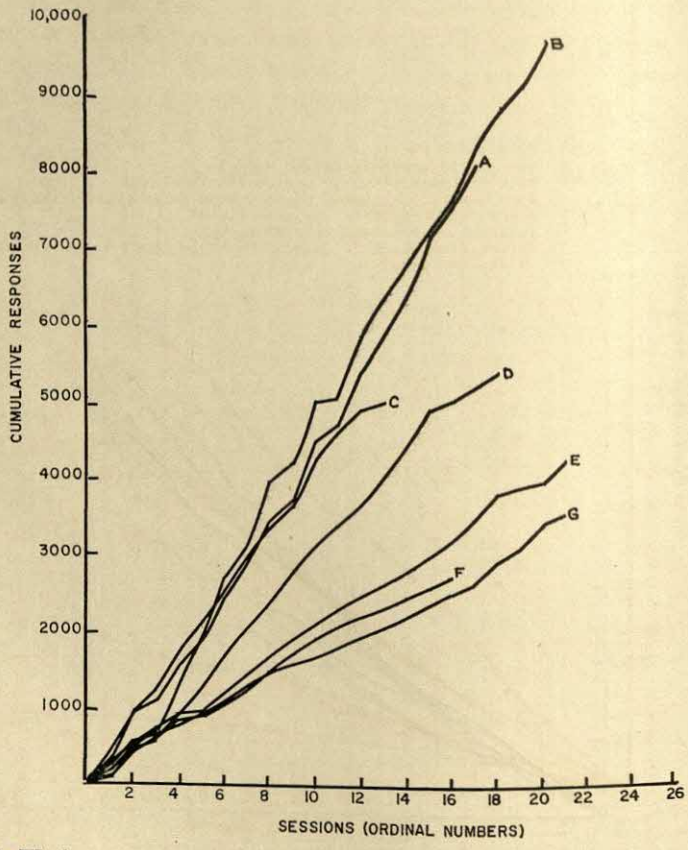


FIG. 7. Work output of retarded readers under conditions of self-selection, self-pacing, and immediate feedback on performance quantity (Class II).

regarding talking or disturbing others were in effect. At the end of each class, correct responses were counted, in the presence of each child, and he charted his total "points" on a daily bar graph and a weekly cumulative bar graph. No comment was made by the teacher.

Results. Work rates for Groups I and II are depicted in Figures 6 and 7. With the exception of four *S*s in Group II (Figure 7), the response curves are remarkably similar to those of Experiment 2. The exceptions are *S* G (IQ, 69) whose rate is slower than the others, but similar in configuration, and *S*s A, B, and C. A and B appeared to be competing with each other. *S* C missed the second week due to a family vacation and worked very rapidly when he returned.

Discussion. Observers reported changes in classroom behavior similar

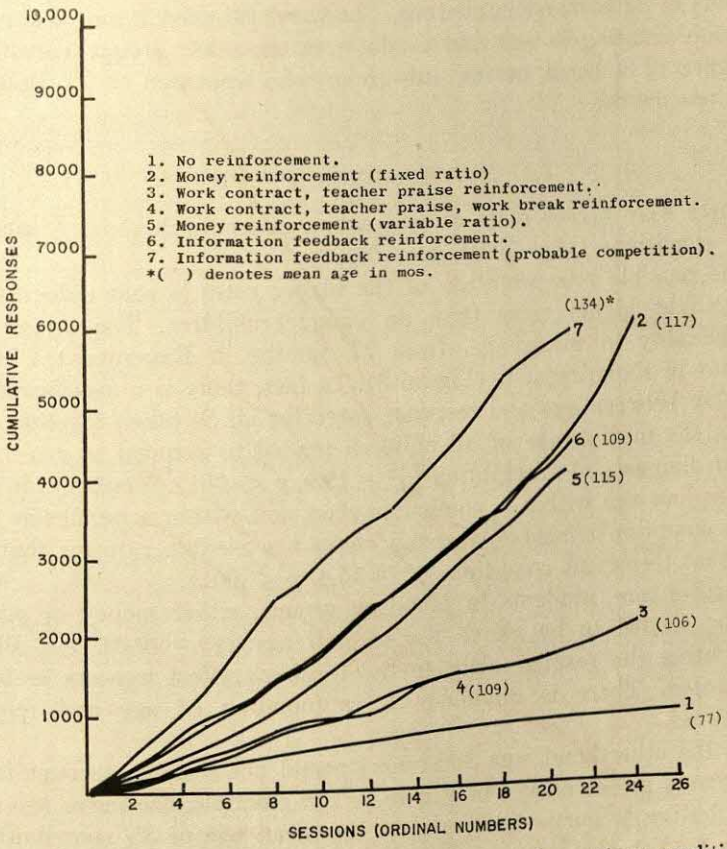


FIG. 8. Median work output of retarded readers under various conditions of reinforcement.

to those reported in Experiment 2. It would appear, on the basis of similarity of work outputs, that information feedback in the form of "points" is fully as effective as the monetary consequence in maintaining task behavior.

SUMMARY OF RESULTS

The results of the six experiments are summarized in Figure 8 by plotting the median work output each day for each class. (Classes 6 and 7 were part of Experiment 6.) Experiment 1, no reinforcement, reflects the lowest work output; 3 and 4 (work contract with teacher praise and teacher praise plus work break) are next in order. These conditions are probably most similar to normative classroom conditions in the schools. Experiments 2, 5, and 6 (money or information feedback) appear similar in level of output and in contour. The curve labelled 7 was produced by a group differing in age and grade from the other groups and showing evidence of a three person sub-group who appeared to be competing with one another.

DISCUSSION

There were higher rates of responding in feedback than in non-feedback conditions ($\chi^2 = 10.04$, $p < .01$).

One possible interpretation of the higher rates is that older children respond at higher rates than do younger children. The groups vary considerably in mean age, from 77 months in Experiment 1 to 134 months in Experiment 6 (Figure 8). In fact, there is a significant relationship between age and response rate: for all *S*s taken together, those above the median age of all children tended to respond at rates above the median rate of all children ($\chi^2 = 18.9$, $p < .001$). Similarly, *S*s above the median age within a particular class and within a particular treatment condition tended to respond above the median rate for that class and that treatment condition ($\chi^2 = 15.4$, $p < .001$).

Furthermore, students in feedback groups, either money or progress charts, tended to be above the median age ($\chi^2 = 10.04$, $p < .01$), so attributing the results solely to the feedback effect appears to be unwarranted. There is obviously a confounding of age and feedback conditions.

On the other hand, age differences would not seem to account for the differences in the curves in Figure 8. For example, the mean age of *S*'s contributing to curve 4 is equal to the mean age of *S*'s contributing to curve 6. Yet the latter curve is consistently higher and accelerating whereas the former curve is lower and, if anything, decelerating. And

curve 5 is lower than curve 6 (younger *S*'s) and lower than curve 2 (older *S*'s).

The problem addressed in these studies is that of maintaining task relevant behavior. Prior clinical work with children in the age range represented here had produced patterns of performance similar to those in Figure 1: deceleration and periods of no task relevant performance.

Examination of all the curves of individual *S*'s reveals that, with performance feedback and/or monetary consequences, performance is maintained or accelerated regardless of age. With teacher praise and/or no performance feedback, performance is not maintained ($\chi^2 = 29.7$, $p < .001$) regardless of age.

Similarly, examination of the curves of *S*'s in Figure 3 reveals that a change to a monetary consequence is followed by an increase in task relevant performance ($\chi^2 = 5.5$, $p < .05$). The increase in rate during the several days after the introduction of money is also significant ($\chi^2 = 8.5$, $p < .01$). Since these curves accelerate and since those for individual *S*'s in Figures 5, 6, and 7 are also either maintained or accelerated, we conclude that it is appropriate to attribute the major results to the presence of performance feedback rather than to the age of the *S*'s.

There were monetary consequences throughout Experiment 5 except for the last three sessions. Removal of the money was followed by a range of effects and was not continued long enough to determine whether or not the performance would continue without it. However, the children were already routinely graphing their results before counting their money. As shown by Experiment 6, graphing is a sufficient condition for maintaining task behavior.

SUMMARY

The work output of retarded readers using a programmed literacy curriculum was observed under various conditions of reinforcement in a controlled classroom. Task behavior was followed by conditions of no consequence, teacher praise, a work-break consequence, a monetary consequence and feedback on amount of work (relative to previous work). Extinction of task behavior tended to occur under conditions of no consequence and of teacher praise. Conditions of monetary consequence and of feedback on progress resulted in high, sustained rates of work behavior.

CONCLUSION

The solution to the problem of maintaining task behavior was found to be a simple one. Providing retarded readers with feedback on progress was sufficient to keep them performing at high rates. In effect, they were

trained to perform desired classroom behavior by an arbitrary, but systematic, point system. The points replaced money within and between experimental groups with no appreciable change in task-relevant behaviors.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 63–73 (1969)

The Effect of Serial Structure on Children's and Adults' Paired-Associate Learning[1]

FRANCIS J. DI VESTA AND MARTIN E. BERNSTEIN
The Pennsylvania State University

Performance was compared of children's and adults' learning of a paired-associate task in which response terms formed an oppositional serial structure. The hypothesis that end items in the structure served as anchors was supported by the finding that they were learned prior to intermediate items even though the word-pairs were presented in random order. The scaling of items in semantic space was related to the obtained serial-position effect. When idiosyncratic structures were considered the serial learning curve was very closely approximated. Children's use of the structure paralleled those of adults in the learning of all but one of five lists. The results suggest that structuring of the task by *S* involves *elaborative activities* in which the materials are organized in meaningful fashion and *selective processes* that determine the order in which items will be learned.

A unique procedure for demonstrating an interesting effect of structure on learning was introduced by DeSoto and Bosley in 1962. They found that presenting elements of a set of words with a *bounded* serial structure (Freshmen, Sophomores, Juniors, Seniors) as response terms in a paired-associate learning task determined the ease of learning a given response term. The rapidity with which a response was learned was a function of its position in the serial structure even though the paired-associates were presented in random order.

Other studies (Pollio, 1966a; Pollio and Draper, 1966; Pollio, 1968; Pollio and Deitchman, 1964) extended this procedure to include the structures involved in the progression of natural numbers and in sets of words with oppositional serial structures such as Beautiful, Pretty, Fair,

[1]This work was supported by research grant HD-02025 from the National Institute of Child Health and Human Development, U. S. Public Health Service. The authors are indebted to Andrew J. Chatlas, Director of Elementary Education in the Bellefonte Area School District, Ott B. Kantner, Principal of the Bellefonte Elementary School, Robert E. Davy, Principal of the Pleasant Gap Elementary School, Sara Moyer, Principal of the Sixth Street Junior High, Philipsburg, and to the fifth-grade teachers Ruth Jones, Christina Poorbaugh, Stanley Rogers, Marjorie Sholl, and Yvonne Smith of Bellefonte School as well as John Shawley of Pleasant Gap for their cordial cooperation, and that of their pupils, in the conduct of this study.

Homely, and Ugly. An analysis of the latter set of words suggested that the order implied within the set was correlated with locations of the words on the evaluative dimension of semantic space (Osgood, Suci, and Tannenbaum, 1957). The words were also similar in the sense that they all refer to values related to the denotative attribute of beauty.

The present series of experiments extended the results of the earlier studies by comparing children's and adult's performance on paired-associate tasks in which there was an underlying oppositional serial structure unifying the response terms. It was reasoned that the serial position effect, in which the end items are learned more easily than the intermediary items, would be attained with college *Ss* to a significantly greater extent than for fifth grade *Ss* on the assumption that college *Ss* would have overlearned the underlying structure implied by the items. In one experiment, the treatments employed by Pollio (1966a) were replicated with children as well as with adult *Ss*. In another experiment response terms were selected which were ordered on the evaluative dimension but which were denotatively dissimilar. The third experiment investigated the relationship of self-imposed structures to the learning of a task with a single underlying oppositional connotative structure but with more than one underlying denotative dimension. Variations within each of the first two experiments were accomplished by replications with different lists.

METHOD

Subjects. The children were 63 fifth-grade pupils from a school district within the vicinity of the University Park campus of The Pennsylvania State University. The college student *Ss* were 63 undergraduate male and female volunteers from the introductory educational psychology course at The Pennsylvania State University. The college *Ss* received course credit for participation in the experiment. None of the *Ss* had participated previously in a verbal learning study. There were 15 children and 15 college students in Exp. I with List I. In each of the remaining treatments there were 12 *Ss* in a group.

Materials. The task materials were simply memory drum tapes separately prepared for each of the experimental conditions. On each tape were four presentations of 15 pairs of words. The stimulus terms were 15 three-letter girl's names such as Nan, Sue, Bab, Lil, Amy, Roz, Flo, Kim, Dot, Eve, Peg, Viv, Ida, Joy, and Cyd. Each name was randomly paired with one of the five response terms in whatever serial order structure was being manipulated for a given experiment. Thus, for example, in the first experiment to be described below the serial structure was Beautiful, Pretty, Fair, Homely, Ugly. The matching procedure resulted in pairing each adjective, as a response term, with each of three different names, as

stimulus terms. The four presentations of the 15 pairs were in different orders as determined by reference to a table of random digits. The five response terms for each experiment were assumed to represent five points along the continuum of the underlying semantic serial-order structure. The specific response terms and their semantic differential ratings will be given in the description of each experiment. A given set of words was used with both adults and children. Normative data regarding the connotative meanings (semantic differential ratings) of words for children were obtained from the Di Vesta and Walls (1969) norms and those for adults were obtained from the Jenkins, Russell, and Suci (1958) or the Gerow and Pollio (1965) norms. If the norms were not available from any of these sources, ratings of a given word were obtained locally.

Procedure. The *Ss* were tested individually. In an initial familiarization trial each pair of words was exposed for a 5-second interval, preceded by the instruction that the task was to learn the adjective (response) paired with the name (stimulus). Following the familiarization trial, the anticipation method for paired-associate learning was used. Longer presentation times were employed for children than for adult *Ss*; the rate was 3-second/2-second for children, and 2-second/1-second for adults on a Stowe memory drum. All *Ss* learned to a criterion of one errorless repetition of the entire set of 15 word-pairs.

RESULTS

There were three experiments in which the serial order structure was varied. In addition there were slight modifications in procedure among conditions which will be noted where appropriate.

Experiment I

The first experiment was a replication of one conducted by Pollio (1966a). The response terms were Beautiful, Pretty, Fair, Homely, and Ugly (List I). Their respective semantic differential ratings on the evaluative dimension (as measured on the Good-Bad scale) were 1.93, 3.93, 3.93, 5.43, and 5.78 for college *Ss*; and 1.91, 2.44, 2.57, 4.07, and 5.16 for fifth grade *Ss*.

The mean number of trials taken by *Ss* in each group to learn the three pairs associated with each of the response terms are displayed in Figure 1.[2] Although there is a definite serial-position effect it is marred by the slightly greater ease in learning the middle term than in learning

[2] Note that while we have referred to curves we have chosen to display the data in histogram form. The reason for this is that had curves been used equal spacing on the baseline would imply the words were equidistant from one another; however, in no case was this the actual situation.

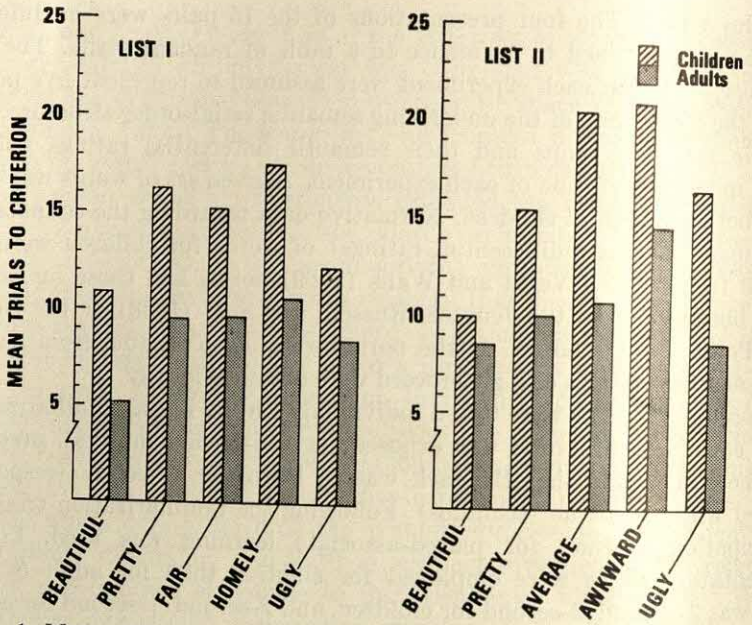


FIG. 1. Mean trials required by children and adults to reach criterion for each element of the oppositional structures in Lists I and II of Exp. I.

each of the adjacent terms. This result contrasts with Pollio's results since he found that the stimuli associated with the response term, Fair, in the intermediate position was considerably more difficult to learn than any of the other terms. However, the present results seem reasonable since they reflect the similarity in the semantic ratings of the two words Pretty and Fair which perhaps leads to interference. Otherwise, the findings compare favorably with those of other studies even to the extent that the item occurring first in the structure (primacy) is learned with greater facility than that item occurring last in the structure (recency). Also to be noted in the data is the degree to which the curves for the two groups parallel one another at every point.

The data for Trials to Criterion for each set of words were analyzed by a mixed analysis of variance with Age as the between variable and response Words as the within variable. This analysis yielded significant main effects due to Age ($F = 10.16$, $df = 1,28$, $p < .01$) and to Words ($F = 6.94$, $df = 4,112$, $p < .01$). The interaction was not significant ($F = 0.62$, $df = 4,112$, $p > .05$).

A second list for Exp. I was devised to provide a serial structure in which the elements were more clearly separated in the children's norms than they were in List I. It was administered in the paired-associate task to *Ss* (children and college students) different from those *Ss* used in the

first task. The list was comprised of the words Beautiful, Pretty, Average, Awkward, Ugly (List II). The norms for adult ratings of these words were 1.93, 3.07, 3.93, 5.14, and 5.78, respectively; the norms for children's ratings were 1.91, 2.44, 3.19, 4.38, and 5.16, respectively.

The results corroborate those found in List I. The definite bow-shaped curve with pronounced primacy and recency effects is displayed in Figure 1. An analysis of variance of Trials to Criterion indicated the effect due to Age was significant ($F = 8.41$, $df = 1{,}22$, $p < .01$) as were the effects due to Words ($F = 12.01$, $df = 4{,}88$, $p < .01$) and to the interaction of Age $\times$ Words ($F = 2.97$, $df = 4{,}88$, $p < .05$).

Experiment II

In the second study an attempt was made to manipulate oppositional structure based on the Good-Bad semantic meaning dimension, but the words themselves did not refer to a similar attribute as did the elements in the Beautiful-Ugly structure. Accordingly, the words Beautiful, Swift, Average, Weak, and Dull (Exp. II-List I) were selected. These words can be seen to represent both the Potency (e.g., Weak) and Activity (e.g., Swift) dimensions of semantic space as well as the Evaluative dimension. The normative semantic ratings on the Good-Bad scale for these words by college students were 1.93, 2.14, 3.07, 5.79, and 6.07; for children the normative ratings, based on scales contributing to the evaluation dimension, were 1.91, 3.29, 2.46, 5.46, and 5.35. The ratings on the Strong-Weak scale were 3.36, 2.21, 3.36, 5.07, and 5.71 for college *Ss* and 3.18, 2.84, 3.00, 6.73, and 5.15 for children. On the Fast-Slow scales the ratings were 3.50, 1.29, 3.50, 5.71, and 6.43 for college *Ss* and 3.97, 2.32, 4.12, 5.26, and 4.73 for children.

The data obtained from the *Ss*' performance on this task are summarized in Figure 2. Again the serial learning curve is approximated to the extent that first and last terms are learned more easily than intermediate terms, except for learning of Average by the group of college *Ss*.

The analysis of variance of these data yielded main effects, on the dependent variable of Trials to Criterion, due to Age ($F = 21.37$, $df = 1{,}22$, $p < .01$) and to Words ($F = 4.03$, $df = 4{,}88$, $p < .01$). The interaction of Age $\times$ Words was not significant ($F = 1.54$, $df = 4{,}88$, $p < .05$).

The second list used in Exp. II, administered to still other groups of *Ss*, was comprised of the items Sweet, Pretty, Usual, Rough, and Dirty (Exp. II-List II). The norms for college students' ratings of these items on the Good-Bad scale were 2.45, 2.45, 2.91, 4.73, and 5.64, whereas children's ratings were 1.48, 2.00, 2.81, 4.94, and 6.34. In this list the oppositional structure was more clearly defined for children than for the

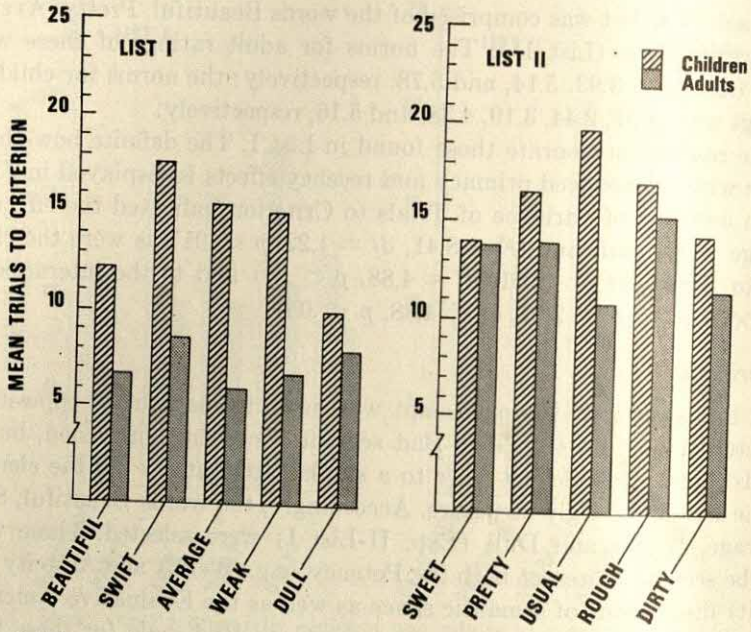


FIG. 2. Mean trials required by children and adults to reach criterion for each element of the structures in Lists I and II of Exp. II.

adult *Ss* and this difference was clearly reflected in the results as may be seen in Figure 2. None of the F ratios for the effects of Age, Words, or the interaction of Age $\times$ Words on Trials to Criterion were significant ($p > .05$) for this list.

Experiment III

The findings from Exp. II indicated clear approximations to the serial structure with children for whom the words were separated by somewhat equal scale distances on the evaluative dimension of the semantic differential; however, the results with adult *Ss* using the same words were not as clear cut. The fluctuations in the curves for these *Ss* were more or less erratic. This finding implied the possibility that some items may stand out over the others, as the word Usual might in the context of the Sweet-Dirty continuum; or, that an idiosyncratic structure was being imposed on the response terms by the adult *Ss*.

Accordingly, the third study was designed to investigate the possibility that under some circumstances *Ss* might impose a structure on a series of response items that corresponds more to their subjective experience than it does to normatively determined structures. It was reasoned that

this structure was an important determiner of whatever learning curves were obtained under such unusual task demands as the presence of several potential structures underlying a series of items. Thus, the learning curves generated in Exp. II could have been influenced, at least in part, by subjectively determined end-anchors and subjectively judged "distances" between elements of the structure.

The words used in the list for Exp. III were Strong, Healthy, Normal, Sickly, and Weak. There are two possible bases for establishing anchors in this series: either Strong-Weak or Healthy-Sickly. The normative adult semantic differential ratings on the Strong-Weak scale for the five terms were 1.14, 1.50, 3.21, 6.79, and 7.00, respectively; and the normative children's ratings were 1.00, 1.34, 2.69, 4.37, and 6.73, respectively. Thus, the terms were clearly separated on this scale for both groups though slightly more so for children than for adults.

In order to determine the self-imposed anchors in this series of items the *Ss* were asked, after completion of the task, to *rank* the five words according to degree of "strength" (most to least) they believed was represented by each word. Trials to Criterion were separately calculated both on the basis of the normative structure and on the basis of the sub-

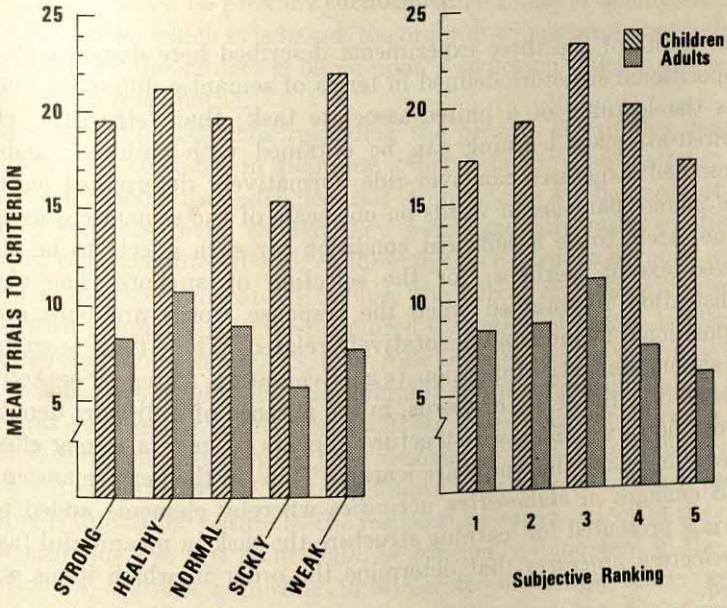


FIG. 3. Mean trials required by children and adults to reach criterion for elements of an ambiguous structure serially ordered according to their semantic differential ratings on the potency scale (left diagram) compared to trials required when elements are ordered according to self-imposed structures (right diagram).

jective judgments. Thus, the items selected as "most strong" and "least strong" by a *S* were assumed to be his anchors; those judged "second most" and "next to least" his second and fourth position terms, respectively, and the remaining item was assumed to be in the middle position.

The performance of *S*s on items ranked according to the normatively determined structure and also according to the self-imposed structure are shown in Figure 3. There it may be seen that, with the serial structure determined by the normative data, a serial learning curve might be approximated if the terms Weak and Sickly were interchanged for the College group, but no amount of "shuffling" of terms would achieve a serial learning curve for the children's responses. However, when the self-imposed structures are considered, the children's and adults' performances yield parallel learning curves that conform almost perfectly to the serial order learning curves. (See Figure 3.)

The analysis of variance of Trials to Criterion based on data for the self-imposed structures yielded significant effects due to Age ($F = 21.88$, $df = 1,22$, $p < .01$) and to Words ($F = 3.55$, $df = 4,88$, $p < .05$). The interaction was not significant ($p > .05$).

DISCUSSION

The results of the three experiments described here demonstrate that an oppositional structure defined in terms of semantic differential ratings affects the learning of a paired-associate task; that "structure" effects in paired-associate learning can be obtained with children; and that self-imposed structures can over-ride normatively determined ones.

The mere separation of words on one scale of the semantic differential does not seem to be a sufficient condition for such effects to be clearly observed except, perhaps, for the selection of anchors. The clearest demonstration is obtained when the response words are both ordered in oppositional fashion *and* denotatively related. Where there is an underlying salient semantic dimension (e.g., Evaluation, Potency, or Activity) represented in the series of words, in the absence of a distinct denotative relationship, a self-imposed structure appears to have a strong effect on the order in which the words are learned. Thus, in the performance of this task *S* engages in *elaborative activities* whereby elements added to the materials presented for learning structure the task in meaningful fashion, and *selective processes* that determine the order in which items will be learned.

As interesting, however, is the finding that the oppositional structure affects children's acquisition of responses in precisely the same manner as it does the adults' learning. With all lists except List II, Exp. II, the

pattern of acquisition by children closely paralleled, point-by-point, the pattern of acquisition by adults. Some of the difference between the performance of the two groups on List II, Exp. II can be explained by the differences in meaning (as defined by semantic differential norms) held by the two groups for the component words of the structure. The data also suggest that the deviations may be the consequence of differences in end-anchors selected by the two groups as basic reference points.

The results of the present study corroborate those of previous studies. They show that learning a series of responses bounded by words on extreme ends of bi-polar scales representing a salient dimension of semantic meaning is similar to paired-associate learning which involves other serial word orders (DeSoto and Bosley, 1962; Pollio and Draper, 1966; Pollio, 1966a). However, it is recognized that words are organized according to many bases (Pollio, 1966a; 1966b). Which principle of organization affects learning depends upon the requirements of the task. For example, in the present study the oppositional words are typically high associates of one another (e.g., Beautiful-Ugly or Strong-Weak) and thus are proximally *related*. Such relationships have been shown to have important effects on mediation, clustering, and other generalization processes because of a structure based on proximity. On the other hand, in the present study, and in studies on which it is based, the order in which words are "selected" for learning by the subject is based upon a structure which suggests these words *oppose* one another.

There are some cautions regarding the present design that might be worthwhile to consider in planning other related investigations. First, there is only moderate assurance that stimulus sets are balanced over the different responses even though they were randomly matched. Especial difficulty was encountered with children in this regard. In our pilot studies, several names were discarded because the effects of their strong associations with popular characters, acquired through such media as television programs and children's books, over-rode the effects of serial order structure. Although such uniquely meaningful words were eliminated, where there was sufficient evidence for doing so, it is conceivable that a given response term may still have been associated with a particularly easy or difficult stimulus set. If it is not known conclusively whether order of learning is due to the nature of the response, as suggested, or to the nature of the stimulus set, this point could be especially important in the interpretation of findings regarding idiosyncratic orderings. More needs to be known about the stimulus attributes cutting across a given response in experiments of this type. While there is the possibility that stimulus factors could have accounted for some of the erratic findings from list to list, the concern seems less important when the degree of replication across

lists, subject samples, and experiments is considered. Nevertheless, the difficulty can be remedied in future investigations by balancing the materials so that each stimulus set is associated with each response an equal number of times. Finally, the effect of two potential anchoring points on self-imposed structures will require at least several more samples of this type of response set before the conclusions will have a definite basis for generalization.

Despite the cautionary note, the procedure appears to be sufficiently sensitive to extra-experimental relationships as to warrant further investigation of the effects of such variables (DeSoto and Bosley, 1962; Pollio and Deitchman, 1964). One of the reasons why the exact replication of the bow-shaped curve is not always attained, as in the present study, may be that some words such as those in the Strong-Weak and Beautiful-Ugly oppositional series are ordered differently by men and women or by boys and girls. This, too, suggests that differences among groups might be hypothesized to be related specifically to the learning of structures peculiarly associated with particular groups. Thus, Beautiful-Ugly might provide more definite anchors for women than for men but Strong-Weak might be more structured for men than for women. Similarly, whether men's or women's names are associated with one or the other of the structures might be found to interact with the kind of structure used (see, e.g., Pollio and Draper, 1966). The consideration of differences among groups suggest, too, that this procedure might also be useful for investigations related to the cognitive functioning of lower- and middle-class children, of children with different ethnic origins, and the like.

REFERENCES

DeSoto, C. B., and Bosley, J. J. The cognitive structure of a social structure. *Journal of Abnormal and Social Psychology*, 1962, **64**, 303–307.

Di Vesta, F. J., and Walls, R. T. A factor analysis of the semantic attributes of 485 words and some relationships to the conceptual behavior of fifth grade children. Studies in Verbal Processes, Report No. 15, January, 1969, Pennsylvania State University, Grant HD-02025, U. S. Public Health Service.

Gerow, J. R., III, and Pollio, H. R. Word association, frequency of occurrence, and semantic differential norms for 360 stimulus words. Technical Report No. 1, April, 1965, University of Tennessee, Grant MH 08903, U. S. Public Health Service.

Jenkins, J. J., Russell, W. A., and Suci, G. J. An atlas of semantic profiles for 360 words. *American Journal of Psychology*, 1958, **71**, 688–699.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. *The measurement of meaning*. Urbana, Illinois: University of Illinois Press, 1957.

Pollio, H. R. Associative structure and verbal behavior. Pp. 37–66. In T. R. Dixon and D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, New Jersey: Prentice-Hall, 1968.

POLLIO, H. R. Oppositional serial structures and paired-associate learning. *Psychological Reports,* 1966, **19,** 643–647. (a)

POLLIO, H. R. *The structural basis of word-association behavior.* The Hague: Mouton, 1966. (b)

POLLIO, H. R., AND DEITCHMAN, R. The activational characteristics of a serial cognitive structure on paired-associate learning. Unpublished mimeographed report, University of Tennessee, 1964.

POLLIO, H. R., AND DRAPER, D. O. The effect of a serial cognitive structure on paired-associate learning. *Journal of Verbal Learning and Verbal Behavior,* 1966, **4,** 301–308.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 74–81 (1969)

Preschool Children's Ratings of Familiarized and Nonfamiliarized Visual Stimuli

GORDON N. CANTOR AND SUNNAN K. KUBOSE[1]
Institute of Child Behavior and Development, University of Iowa

Two experiments are described in which preschool-aged *Ss* first observed repeated projections of six stimuli (white on black, low meaningful figures) in a familiarization phase and then rated these and six previously unseen stimuli in a test phase. The test phase for Experiment I entailed single stimulus presentations, with *S* responding to each by saying, "I like it" or "I don't like it." In Experiment II, familiarized and nonfamiliarized stimuli were presented in pair-wise fashion in the test phase, *S* being asked to point to the one he "liked best" in each case. The results from both experiments indicate the *Ss* showed a significant preference for the nonfamiliarized over the familiarized stimuli. These findings are discussed in relation to: (a) those of earlier related child studies; (b) a thesis recently proposed by Zajonc (1968) to the effect that "mere exposure" to stimuli enhances attitudes toward them; and (c) the position taken by Fowler (1967) that familiarized stimuli are a source of tedium and stimulus change may constitute a reinforcing state of affairs.

In a recently published monograph, Zajonc (1968) discusses evidence supporting the thesis that, ". . . mere repeated exposure of the individual to a stimulus is a sufficient condition for the *enhancement* (italics ours) of his attitude toward it." By "mere exposure" Zajonc means, ". . . a condition which just makes the given stimulus accessible to the individual's perception" (Zajonc, 1968, p. 1). Support for this thesis is impressively documented by reference to varied kinds of evidence, including correlations found between affective connotations of words and word frequency in the natural language, the effects of experimentally manipulated frequency on affective responses to symbols or their referents, etc.

Data gathered on child *Ss* in our laboratories and elsewhere over the past several years provide a consistent and puzzling contrast to the sorts of findings cited by Zajonc. For example, Cantor and Cantor (1964a, 1964b, 1966) have found that kindergarten-aged children given the opportunity to project familiarized and nonfamiliarized visual stimuli project the nonfamiliarized material for longer periods of time. Such

[1] We are indebted to Mrs. Mary Jane Oltmans, who gathered the data for Experiment II. The work reported in this paper was conducted while the second author was supported by NICHD Training Grant T1-HD-113.

results have suggested the conclusion that, given the particular stimuli and the familiarization and test procedures involved, *Ss* "prefer" the nonfamiliarized to the familiarized stimuli.

We have been interested more recently in obtaining supportive evidence for such a conclusion by asking child *Ss* to provide "like-dislike" ratings of familiarized and nonfamiliarized stimuli. The desirability of tapping preferences via use of different response systems is attested to by the recently reported data of Kaess and Weir (1968), who found that 29- to 66-month-old children's preference functions for stimuli varying in complexity level differed when stated preferences and observing times were used as dependent variables.

In an initial study (Cantor, 1968), 5th- and 6th-graders evaluated familiarized and nonfamiliarized Welsh figures (Welsh, 1959) on a "like-dislike" continuum. It was found that these *Ss* gave significantly more favorable ratings to nonfamiliarized than to familiarized stimuli. The present paper reports two follow-up rating experiments that entailed use of preschool-aged *Ss*. Since the *projection* data cited above were gathered on kindergarteners and the just mentioned *rating* data were obtained on 5th- and 6th-graders, it was considered desirable to gather rating data on *Ss* more comparable in age to those serving in the projection studies than was the case in the initial rating study. We believe the results of these most recent experiments are of interest because: (1) they suggest that, under certain circumstances, observing time and rating measures do produce comparable sorts of findings; (2) the data add to a growing body of evidence that conflicts with the Zajonc thesis; and (3) such findings bear rather directly on current theorizing (e.g., Fowler, 1967) that attributes tedium properties to familiarized stimuli and reinforcement properties to stimulus change.

EXPERIMENT I

In this experiment, *Ss* viewed six 4-second projections of each of six stimuli (white on black, low meaningful figures) and then rated these and also six others (not previously seen) by responding "I like it" or "I don't like it." Thus, the test phase in this case entailed simple verbal responding on the part of the *Ss* and single-stimulus presentations.

Method

Subjects

The *Ss* were 52 children (28 boys, 24 girls) enrolled in the University of Iowa Preschools. Their mean age was 4 years, 11 months, with a range of 4 years, 0 months to 5 years, 8 months.

Apparatus and Materials

The principal pieces of apparatus were a Kodak RA950 Carousel projector, two Hunter timers used to control stimulus and interstimulus intervals, and a rear screen projection box. Twelve figures selected from the Welsh Figure Preference Test (Welsh, 1959), divided into two sets (A and B) of six figures each, were used as stimuli.[2]

Procedure

Each *S* was run individually in a dimly illuminated room. The session began with a *familiarization phase* consisting of six 4-second exposures of each of the six figures in either set A or set B, for a total of 36 stimulus presentations. Half the boys and half the girls were assigned randomly to be familiarized on set A, the remaining *S*s being familiarized on set B. For a given familiarization group, each of the six stimuli was presented once within successive blocks of six trials. A different random order was used for each of the six blocks, these orders being the same across all *S*s in that familiarization group. Interstimulus intervals lasting 2.5 seconds entailed projections of a blank slide. In the familiarization phase, *S* was simply told to attend to the projection screen.

At the start of the *test phase,* which followed immediately after the familiarization phase, the following instructions were read to *S:*

> Now when I show you a picture, I want you to look at it carefully until it goes off. After it goes off, tell me if you like it or if you don't like it. I have looked at these pictures before and I like some of them but I don't like some of them. You look at the pictures and after each one goes off, say "I like it" if you like it or say "I don't like it" if you don't like it. (If *S* did not respond following a stimulus presentation, *E* asked, "How about that one?")[3]

Each of the 12 figures (six familiarized and six nonfamiliarized) was projected once for a 4-second duration during the test phase. The stimuli were grouped into six blocks of two each, with one familiarized and one nonfamiliarized stimulus being randomly assigned to each block. The

[2] Permission to reproduce these figures was obtained from George S. Welsh and Consulting Psychologists Press, Inc., Palo Alto, California. The 12 stimuli used in the present study were selected from those reproduced in Cantor and Cantor (1964a). Those numbered 1, 4, 6, 7, 9, and 10 in the earlier paper composed set A in the present study; those numbered 12, 14, 16, 17, 19, and 20 composed set B.

[3] In the earlier rating study (Cantor, 1968), 5th- and 6th-graders evaluated stimuli by marking points along a 12.8-cm line having as anchoring terms the words "strongly dislike" and "strongly like." This task was judged to be inapplicable in the case of preschoolers—hence the use of the simpler rating task described above in Experiment I and the subsequently described task utilized in Experiment II.

order of occurrence within each block was also randomly determined. Four different such semi-random stimulus orderings were used, a fourth of the *Ss* being assigned to each of the sequences.

The length of the interstimulus interval in the test phase depended on the latency of *S*'s evaluative response and/or the time required by *E* to record *S*'s response, select the next figure to be presented, and reset a timer (approximately 5 seconds). Despite the instructions, *Ss* frequently responded during the 4-second exposure duration. In these cases, *E* recorded the response when it occurred and, following stimulus offset, performed the remaining two steps; thus, the interstimulus interval in these instances depended only on the time required for the necessary procedural steps.

At the completion of the test phase, *S* selected a prize from a display of dime store toys and returned with it to his classroom.

RESULTS

The criterion measure was the number of times *S* responded with "I like it" to the familiarized and to the nonfamiliarized stimuli. The possible range for a given *S* was 0–6 for each of the two stimulus categories. The average number of times the 52 *Ss* responded "I like it" were as follows: (a) for the *familiarized* stimuli—$\bar{X} = 3.92$ ($SD = 1.50$); and (b) for the *nonfamiliarized* stimuli—$\bar{X} = 4.77$ (SD = 1.28).

Note that both of these values exceed the average to be expected (i.e., $\bar{X} = 3.00$) had the *Ss* been responding on a random basis. Applications of the appropriate t test revealed that in both instances the *Ss* responded "I like it" significantly more frequently than would be expected by chance (t for familiarized stimuli $= 4.42$, $df = 51$, $p < .001$; t for nonfamiliarized stimuli $= 9.97$, $df = 51$, $p < .001$).

However, when a related t test was run comparing the numbers of times the *Ss* responded favorably to the two stimulus types, it was found that "I like it" responses occurred significantly more often to the nonfamiliarized than to the familiarized stimuli ($t = 3.35$, $df = 51$, $p < .01$). Thus, despite a general tendency to engage in "yea saying," the *Ss* did respond so as to express a greater liking (as defined by the rating method used) for the nonfamiliarized stimuli.

EXPERIMENT II

In the second experiment, *Ss* viewed six 5-second projections of each of six stimuli. Subsequently, each of the six familiarized stimuli was paired with a different one of six previously unseen stimuli. The *S* responded to each paired presentation by pointing to the stimulus he "liked best." Hence, the test phase in Experiment II entailed paired-presentations of stimuli and a simple motor response on *S*'s part.

Method

Subjects

The *Ss* were 36 children (16 boys, 20 girls) enrolled in the University of Iowa Preschools. They ranged in age from 4 years, 2 months to 5 years, 2 months, with a mean of 4 years, 7 months. None of these children served as *Ss* in Experiment I.

Apparatus and Materials

In addition to the necessary timing equipment, this experiment entailed use of two Kodak RA950 Carousel projectors and two rear-screen projection boxes. The same Welsh figures used in Experiment I, divided as before into two sets, composed the stimulus materials.

Procedure

The *S* was told before entering the experimental room that he would win a toy if he watched "carefully" during the game he was to play. Half the *Ss* (9 boys, 9 girls) were familiarized on stimulus set A, the remaining *Ss* (7 boys, 11 girls) being familiarized on set B. The *familiarization phase* was administered exactly as in Experiment I (including use of one projector and one screen, the latter located directly in front of *S*), except that 5-second stimulus-on and 3-second stimulus-off intervals were used and *S* was told after every six projections, "Be sure to watch carefully" or "Keep on watching carefully—that's good." Immediately following the familiarization phase, *E* rearranged the situation by introducing the second projection box so that the two boxes were placed symmetrically before *S*, with the two projectors being located in appropriate positions behind the boxes. Each projector was loaded with slides involving all 12 stimuli, so that any figure could be projected on either screen.

To begin the *test phase*, *S* was then told, "Now I'm going to show you two pictures at a time. When the pictures come on, look at *both* of them carefully, and then point to the one you like best." The six paired presentations were then given, a trial beginning with the simultaneous appearance on the screens of a particular stimulus pair. After *S* pointed to one of the stimuli, *E* switched the projectors to blank slides, recorded *S*'s response, and then presented the next stimulus pair. Following the sixth presentation, *S* chose a reward from a pool of dime store toys and returned to his classroom.

The composition of the paired presentations in the test phase was determined as follows. Consider the set A stimuli as numbered 1–6, the set B stimuli as 7–12. Six different patterns of familiar stimulus-non-

familiar stimulus pairings were set up (e.g., one pattern was 1–7, 2–8, . . . , 6–12). In no case was a particular pairing repeated across the six pairing patterns. For each of the six pairing patterns, six different orders of paired presentations were then set up. Three *Ss* from the group familiarized on stimulus set A were assigned to a different one of the six orders for a given pairing pattern and three *Ss* from the group familiarized on stimulus set B were assigned to the remaining three orders for the pairing pattern in question. Similar assignments were made to the remaining pairing pattern-order combinations, with the consequence that there was a total of 36 pairing pattern-order combinations, one for each of the 36 *Ss*. The right-left spatial arrangement of the stimuli within pairs was randomly determined for each *S* separately, with the restriction that the familiar and nonfamiliar stimuli occurred equally often on the right and left in the six paired presentations given each *S*.

RESULTS

The dependent variable was the number of times *S* pointed to the *nonfamiliarized* stimulus as the preferred member of a pair, the possible range of scores being 0–6. The mean across all 36 *Ss* was 4.69 ($SD = 1.39$), indicating a preference for the nonfamiliarized stimuli. Testing of the obtained mean against what would be expected by chance (i.e., $\bar{X} = 3.00$) produced a $t = 7.29$ ($df = 35$, $p < .001$). Twenty-nine of the 36 *Ss* had a score of 4 or higher. Because 6 of the 7 *Ss* who deviated from this preference tendency were in the group familiarized on stimulus set B, it was decided to examine the data separately for the two counterbalancing groups. The group familiarized on set A had a mean of 5.11 ($SD = 1.02$), whereas the mean for those *Ss* familiarized on set B was 4.28 ($SD = 1.60$). Both of these means deviate significantly from 3.00 ($t = 8.79$, $df = 17$, $p < .001$; $t = 3.37$, $df = 17$, $p < .01$).[4] While the two groups differed significantly from each other in the magnitude of their variances ($F = 2.45$, $df = 17/17$, $p < .05$), the results of neither a t test nor a median test suggested that a difference in central tendency between the two groups was significant.[5]

[4] These two t values must be interpreted guardedly, in view of the skewed nature of the data and the small N's involved here. Note, however, that the t test is quite appropriate when applied to all 36 scores, as was the case with the t test result reported earlier.

[5] An additional 36 *Ss* were run in conjunction with Experiment II, with the procedures being identical to those used in the main experiment, save for the fact that *S* was told in the test phase to *point to the stimulus he had seen* during the familiarization phase. Surprisingly, the accuracy of recall ($\bar{X} = 4.44$, $SD = 2.01$) was smaller in magnitude (but not significantly so—$t = .61$, $df = 70$) than the preference

DISCUSSION

The results of the two experiments reported in this paper are consistent with those of the earlier rating study (Cantor, 1968) and three projection-time studies (Cantor & Cantor, 1964a, 1964b, 1966) cited previously, as well as with those reported by others using child *Ss* but different sorts of experimental paradigms (e.g., Odom, 1964; Harris, 1965; Mendel, 1965; Gullickson, 1966; Endsley, 1967).

The picture emerging from all of this work—that of the child's preference for nonfamiliar as opposed to familiar stimuli—appears to collide head-on with the thesis of the Zajonc (1968) monograph. In reflecting on this discrepancy, one's attention is immediately directed to the fact that all of the empirical evidence cited by Zajonc was gathered on the high school or college student. Perhaps developmental factors are of crucial relevance here.

Were Zajonc's conclusion limited to the sphere of the word frequency-attitude relationship (and a substantial portion of his monograph has such a focus), the discrepancy might not be considered a compelling one. It may well be that the consideration of differential numbers of exposures (but all numbering in the thousands or tens of thousands or more) to word stimuli has little or nothing in common with that of the exposure of children to small numbers of presentations of geometric forms, toys, etc. But Zajonc discusses experimental studies that entailed small numbers of exposures and use of non-word stimuli (e.g., Chinese characters, pictures of men's faces). Thus, there is a generality to the data Zajonc cites that leaves one with the definite impression that the *Ss* in the child studies mentioned above should have been expected to prefer the familiarized to the nonfamiliarized stimuli presented to them.

In concluding his monograph, Zajonc urges that, ". . . future research must, in particular, concentrate on the effects of large frequencies of exposure, on duration of exposure, on interexposure intervals, and on many other similar parameters of mere exposure" (Zajonc, 1968, pp. 23-24). To this list one might add type of exposure, variation in frequencies at the low end of the scale, and, so it would seem, the age of the *S*.

It is probably worth noting that the theorizing and animal data recently provided by Fowler (1967) also conflict with the Zajonc thesis. Fowler suggests that familiar stimuli are tedium producers, whereas stimulus change within certain boundary conditions constitutes a reinforcing state of affairs. It would appear that the available child data fit better with the notions of Fowler than with those of Zajonc.

---

value ($\bar{X} = 4.69$) obtained in the main experiment. However, the recall score mean is significantly greater than would be expected by chance ($t = 4.30$, $df = 35$, $p < .001$).

REFERENCES

CANTOR, G. N. Children's "like-dislike" ratings of familiarized and nonfamiliarized visual stimuli. *Journal of Experimental Child Psychology*, 1968, **6**, 651-657.

CANTOR, JOAN H., AND CANTOR, G. N. Observing behavior in children as a function of stimulus novelty. *Child Development*, 1964a, **35**, 119-128.

CANTOR, JOAN H., AND CANTOR, G. N. Children's observing behavior as related to amount and recency of stimulus familiarization. *Journal of Experimental Child Psychology*, 1964b, **1**, 241-247.

CANTOR, JOAN H., AND CANTOR, G. N. Functions relating children's observing behavior to amount and recency of stimulus familiarization. *Journal of Experimental Psychology*, 1966, **72**, 859-863.

ENDSLEY, R. C. Effects of differential prior exposure on preschool children's subsequent choice of novel stimuli. *Psychonomic Science*, 1967, **7**, 411-412.

FOWLER, H. Satiation and curiosity: Constructs for a drive and incentive-motivational theory of exploration. In K. W. Spence and Janet T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. 1. New York: Academic Press, 1967.

GULLICKSON, G. R. A note on children's selection of novel auditory stimuli. *Journal of Experimental Child Psychology*, 1966, **4**, 158-162.

HARRIS, L. The effects of relative novelty on children's choice behavior. *Journal of Experimental Child Psychology*, 1965, **2**, 297-305.

KAESS, D. W., AND WEIR, M. W. Response technique as a factor in determining children's preference for complexity. *Psychonomic Science*, 1968, **11**, 367-368.

MENDEL, GISELA. Children's preferences for differing degrees of novelty. *Child Development*, 1965, **36**, 453-465.

OSOM, R. D. Effects of auditory and visual stimulus deprivation and satiation on children's performance in an operant task. *Journal of Experimental Child Psychology*, 1964, **1**, 16-25.

WELSH, G. S. *Welsh figure preference test* (Research edition). Consulting Psychologists Press, 1959.

ZAJONC, R. B. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph Supplement*, 1968, **9**, No. 2, Part 2.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 82-91 (1969)

The Relevance of an Adequate Concept of "Bigger" for Investigations of Size Conservation: A Methodological Critique

ERNEST A. LUMSDEN
University of North Carolina at Greensboro

AND

J. KENNETH KLING[1]
Florida State University

Having found that the concept of "bigger" for young children is disproportionately weighted for verticality (Lumsden and Poteat, 1968), the hypothesis tested here is that this inadequate concept has been a confounding source in investigations of size conservation which have utilized the question, "Is one bigger than the other?" Training directed toward the learning of a multidimensional concept of "bigger" was administered to 20 children ranging in age from 5½ to 7½ yrs. A standard size-conservation task was given to this experimental group and a control group of 20 comparable children of the same age range. The difference between the incidence of size-conserving responses of the trained and untrained groups was significant beyond the .01 level. Furthermore, a posttest of the adequacy of the "bigger" concept revealed a difference significant beyond the .001 level. These data point to the necessity of assuring that conservation questions be free of preconceptual concepts not yet attained by the youngest of the Ss to whom the question will be addressed.

The investigation of conservation of size entails presenting the child with two objects of the same shape and size and, while he is looking on, transforming the shape of one of the two objects. At this point a question that is sometimes asked is, "Is one *bigger* than the other?"

Some investigators require only that a negative reply be given in response to this inquiry as sufficient evidence of conservation of size, while others require further that the child give reasons for his negative response that reflect the psychological operations (reversibility, compensation, and logical necessity) presumed to underlie the conservation

[1] This experiment was carried out by the junior author who is now in the doctoral program at Florida State University.

of this attribute (Gruen, 1966). We wish to emphasize, however, that all investigators interpret a reply of "yes" to the initial conservation question as definitely indicating a lack of conservation. This paper attempts to demonstrate the error inherent in this practice when the conservation question takes the form of "Is one *bigger* than the other?"

Lumsden and Poteat (1968) reported an investigation revealing an extremely disproportionate weighting of the vertical dimension in the concept of "bigger" in preschool children. On the basis of this finding, they suggested that the conservation question utilizing the comparative adjective, "bigger," permits a confounding of the adequacy of the child's concept of this word with the operations which the conservation task is intended to disclose. Their findings imply that a young child may indicate that the altered object was "bigger" simply because it had been made taller. Such a child possibly could have given evidence of conservation of size had the question not employed the word "bigger." It would appear, therefore, that the use of this question in an attempt to determine the age at which children are capable of size conservation might assure that it is not expressed until after their concept of "bigger" is congruent with that of the investigator.

Conservation as determined by the operations listed above has been referred to in the literature as conservation of size, amount, quantity, substance, matter, as well as conservation of material. In the interest of more definitive communication, perhaps it should be recognized that the conservation of size, however, is a special instance of a larger category of conservation for which all of the other terms are equally adequate. These terms encompass alternative transforming operations as well as the simple transformation of shape of one of two identical forms of equal amounts of material with which we are here concerned. Examples of other transformations are (a) dividing one of two identical forms of material into two or more parts; (b) dissolving one of two equal cubes of sugar in water; and (c) "popping" one of two equal amounts of pop corn (an operation which effects a change in density while leaving the amount of matter unchanged). Therefore, the thesis of the investigation to be reported here is pertinent only to one of the questions employed in determining the conservation of size since the word "bigger" would hardly be employed in the conservation question following any of the other transformations of amount mentioned above.

The following general design is offered as a test of the relevance of the concept of "bigger" for evidencing size conservation in response to the question "Is one bigger than the other?" Two groups are randomly formed from a group of young children not selected in regard to the adequacy of their concept of "bigger" and one group is trained in a

multidimensional concept of "bigger." It is hypothesized that to the extent that the training is effective there will be a higher incidence of size-conserving responses in this group than in the untrained group. Since we are not taking the position that an adequate concept of "bigger" is the only condition necessary for the expression of size conservation in this situation, there may well be those who do evidence an adequate concept but fail to evidence size conservation. Presumably this condition would be more prevalent among the younger *Ss*.

METHOD

Subjects. The *Ss* were 40 children ranging in age from 5½ to 7½ years: 20 were kindergarten pupils from 5½ to 6½ years of age and 20 were first-grade pupils ranging in age from 6½ to 7½ years. As the *Ss* appeared for participation they were alternately assigned to the experimental and control conditions with the restriction that the younger and the older group were equally represented in both conditions and that both sexes were equally represented in both conditions within each age group. Proceeding in this manner, the experimental group consisted of 20 children, 10 ranging from the ages of 5½ to 6½ years and 10 ranging from the ages of 6½ to 7½ years. The same is true for the control group.

Training stimuli and apparatus. In the attempt to expedite experimentally the development of an adequate concept of "bigger" with the experimental *Ss*, 19 gray wooden blocks were utilized ranging in volume from 2 to 36 cubic inches. These blocks were presented to individual *Ss* in pairs on a turntable situated directly in front of the subjects. The turntable was electrically controlled by an automatic timer which allowed it to make one-half of a revolution, requiring 14 seconds. The presentation schedule provided for each pair of objects to be presented twice; when a pair appeared again, laterality was reversed.

For the control group, a tachistoscope, Model D-0950 produced by Polymetric Products was used in an object-recognition task involving six white wooden blocks of equal height and volume, differing considerably in form.

A penny-dispensing device was constructed and utilized as a means of providing reinforcement for the responses of both the experimental and control *Ss*. This black, wooden device permitted the subject to view the penny dispensed by the experimenter as it descended down a zig-zag track to the bottom where the pennies accrued until the end of the training for each *S*. It also permitted the experimenter to withdraw a penny whenever the *S* made an incorrect response. In this manner the experimenter was able to reward and punish the *Ss* appropriately without having to risk the unpredictable consequences of money exchange

with the child and at the same time allow the *S* to see how many pennies he had "won" at any point of the training session or the object-recognition task.

Procedure. Training of experimental *Ss*: The instructions to each of the experimental *Ss* were:

> We are now going to play a game with wooden blocks. When I open the door (*E* pointed to the sliding door in front of the turntable) you will see two blocks going around very slowly on a table like a record player. When they stop going around, I want you to pick up the bigger wooden block and push it through the hole in the box (*E* pointed to the hole in the front of the panel to the left of the turntable). But first, in talking about wooden blocks, "bigger" means the block that has the most wood in it. The "bigger" wooden blocks will not always be the "taller" blocks. They will not always be the "fatter" block. Sometimes the bigger block will be the taller, sometimes it will be the shorter. But always the bigger block is the one with the most wood in it. That's what "bigger" means when we talk about wooden blocks. Instructions regarding the penny-dispensing device were read at this point. Every time you are right and give me the block that is really the bigger one, you will win a penny (*E* activated dispenser lever and a penny rolled down the track). . . . Every time you are right you get a penny (*E* activated the dispenser lever again) . . . but every time you are wrong, you lose a penny (*E* activated the subtraction lever and a penny fell from *S*'s view) . . . every time you are wrong, you lose a penny (*E* activated subtraction lever again).

The training consisted of four progressive stages, requiring that a criterion be met at each of the first three phases before progressing to the next. From Phase I through Phase IV, the pairs of blocks were increasingly more similar in volume until at Phase IV some pairs were utilized of equal volume. The volumetric ratios of the pairs by phase are as follows: Phase I, 1:3; Phase II, 1:2; Phase III, 2:3; Phase IV, 2:3 and 1:1. Phases I, II, and III involved the usage of eight pairs of blocks presented twice in two different sequences; laterality was reversed on the second presentation. Thirteen pairs of blocks were utilized in Phase IV, six pairs having a volumetric ratio of 2:3 randomly interspersed with seven pairs of blocks of equal volume but different vertical dimensions. Throughout all of the training, the smaller block was taller than the larger block as often as the obverse. The equal-volume pairs of Phase IV consisted of blocks of different height. In this way height, *per se*, was rendered irrelevant to an accurate judgment of "bigger." The

criteria of performance for the first three phases were as follows: Phase I, seven correct trials of the last eight trials administered; Phase II, six correct trials of the last eight trials administered; and Phase III, six correct trials of the last eight trials administered. Phase IV did not have a performance criterion requirement. The intended function of Phase IV was to lessen any expectation that one object would always be bigger than the other as was always the situation in the first three phases. Such a set would probably bias the performance of the experimental group in the conservation problem that was to follow. A criterion did not seem appropriate for this phase since, having met the criteria of the first three phases, it was assumed that *Ss* would have already demonstrated a concept of "bigger" sufficient to permit a test of our hypothesis without further requiring them to recognize that two differently shaped objects were precisely the same size. Just prior to Phase IV, *E* read the following instructions.

> Now, we're going to change the game a little bit. So far, one object has always been bigger than the other object. For this next part of the game, sometimes one object will be bigger than the other object, just like the last few times, and sometimes one object won't be bigger than the other object, they'll be the same size. That is, sometimes one will be bigger than the other and sometimes one won't be bigger than the other.
>
> I'll show you two objects and ask you if one is bigger than the other. If you think that one is not bigger than the other say "no." If you do think that one is bigger than the other, say "yes" and pick up the bigger one and put it in the hole in the box just like before. I'll tell you again: if you don't think that one is bigger, say "no" when I ask you. If you do think that one is bigger, say "yes" and pick it up and put it in the hole. Okay?

Please note that during this phase, involving some equal-volume pairs, *E* asked the same question as that to be employed in the conservative problem that follows this phase: "Is one bigger than the other?"

When the *S* made an incorrect choice during any of the four phases, the *E* immediately pointed out the compensatory dimensional relationship between the two objects on which a correct judgment would have been based. The explanations were read from a prepared list of statements appropriate to an incorrect response for each pair of blocks utilized in the training. This verbal feedback was in addition to the penny received for each correct choice and the loss of a penny for each incorrect choice.

Training of control Ss. A comparable reinforcing experience with *E*

was provided for the control group by training intended to be irrelevant both to the development of an adequate concept of bigger and to the size conservation task. The objects utilized with this group were all of the same volume and of the same height, differing only in shape. An object indicated as the standard was tachistoscopically exposed for 20 seconds. Subsequently, the *S* was presented different orientations of the standard object as well as any of five other objects for a duration of 200 milliseconds. His task was to say "yes" if the object being exposed was thought to be the standard to which he had prior exposure and "no" if the object were different from the standard. As was the case in the training of the experimental group, *E* dispensed a penny for correct responses and withdrew a penny for incorrect responses.

Size conservation task. Immediately following the training of experimental *Ss* or the object-recognition task of the control *Ss*, *E* displayed two identical cubes of plastic and read the following instructions to *S*:

> 'I have two cubes of clay here that are exactly the same size, exactly the same size. Now I'm going to change the shape of this one (*E* chose one of the cubes of clay and transformed its shape to a form twice the original height with compensating decreases in the width and depth and placed both objects on the turntable with the major dimension of the altered object oriented vertically. As the table revolved, he asked the conservation question). . . . Is one bigger than the other?

Regardless of *S*'s response to the question "Is one bigger than the other?" the experimenter asked "Why?" The responses to each of these questions were recorded.

The criteria for accepting a response as a size-conserving response were that *S* indicate that the altered object was of the same size as the unaltered one and that he be able to give a logical reason for this condition either in terms of the reversibility of form or by stating the logical necessity that the two objects be equal in size since no material had been added or subtracted in the process of the shape transformation. It is important to note that we did not accept a reason which simply pointed out that although the altered object was much taller than the unaltered one, the width and thickness were commensurately smaller. This is to say that, unlike other investigators, we did not accept a compensatory rationale that did not also include an explicit reference to the prior state of the altered object as evidence that size had been *conserved*. Such a rationale could readily be offered by one who did not even witness the crucial transformation of the shape of the object in question! These restricted criteria operate in such a way as to minimize

the difference in the incidence of conserving responses between our experimental and control groups since, as would be expected from the nature of the training of the experimental *Ss*, four experimental *Ss* gave this unacceptable rationale whereas only two control *Ss* did so. Furthermore, had we not excluded compensatory rationales from those counted as conserving responses, we would have been at least partially vulnerable to the criticism that the training of the experimental *Ss* simply served to create a set for a multidimensional analysis of the terminal dimensions of the objects of the size-conservation task. This training was not basic to either of the rationales which we accepted as evidencing size conservation. We submit that the training experience is relevant only to a full understanding of the conservation question.[2]

Assessment of the concept of "bigger." It was desirable to have knowledge of the adequacy of the concept of "bigger" of control *Ss* as well as the experimental *Ss*. (It seemed judicious to test the concept after the conservation task lest we call attention to verticality.) For this purpose a pair of blocks was utilized which had been employed in Phase IV. These blocks had a volumetric ratio of 2:3. The larger block was of lesser height than the smaller. This test was conducted in the same manner as was the training of the experimental *Ss*. Data were obtained in this manner with the control *Ss* and compared with the data obtained from the experimental *Ss* with this pair of blocks when it had appeared as an integral part of the training session.

RESULTS

The Fisher Exact Probability Test (Hayes, 1963) was utilized in all of the analyses performed on the data of this investigation. However, the probabilities will simply be reported in terms of their relation to the .01 or .001 acceptance levels.

Evaluation of the training of the experimental Ss. The difference between the experimental and the control *Ss* in regard to the frequency of correct responses in the concept-of-bigger test was significant beyond the .001 level. Analyzing the data of the 5½–6½-year-old *Ss* of both groups

[2] At this point, a weight conservation task was administered to all *Ss* to provide a means of assessing the pervasiveness of the effect of training in an adequate concept of "bigger." The major problem did not permit counterbalancing the order of the weight and size conservation tasks. Subsequent to the completion of this experiment, the authors became aware of an experiment demonstrating that the "transfer" between these two tasks is considerable when they are given in immediate succession (Smedlund, 1961). Fortunately, our purpose for including the weight conservation was not pertinent to the thesis of this investigation. For this reason, no further mention will be made of it except to say that the frequency with which weight conservation was evidenced by the experimental and control groups was precisely the same.

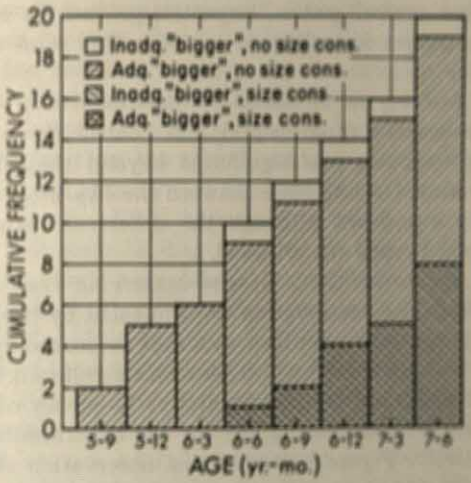


FIG. 1. Cumulative frequency of experimental *Ss* with increasing age in regard to the adequacy of their concept of "bigger" and whether or not they evidenced size conservation.

separately revealed a difference significant beyond the .01 level. The same level of significance was obtained with the data of the 6½–7½-year-old *Ss* as well. Figures 1 and 2 portray the cumulative frequency of *Ss* evidencing an adequate concept of "bigger" with increased age for the

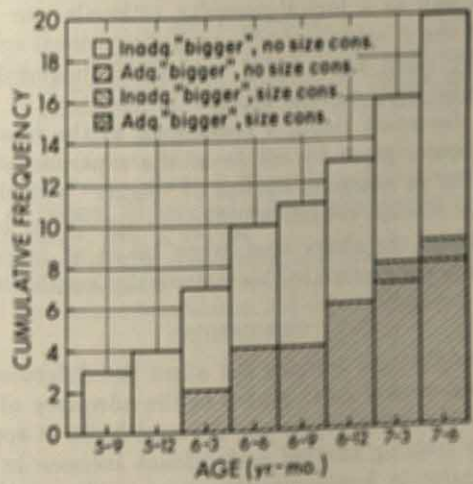


FIG. 2. Cumulative frequency of control *Ss* with increasing age in regard to the adequacy of their concept of "bigger" and whether or not they evidenced size conservation.

experimental and control groups, respectively. Such *Ss* are further divided in these figures according to whether they also evidenced size conservation.

Relation of training to incidence of size-conserving responses. The differences between the experimental and control *Ss* in frequency of size-conserving responses was significant beyond the .01 level. When analyzed separately, the difference between the 5½–6½-year-old *Ss* was not statistically significant although the difference between the older *Ss* was significant beyond the .01 level.

Relation of age to incidence of size-conserving responses. The frequency with which size conservation was indicated by the 5½–6½-year-old *Ss* of both groups was compared with that evidenced by the 6½–7½-year-old *Ss* of both groups. The difference was significant beyond the .01 level. Figures 1 and 2 display the cumulative frequency of *Ss* evidencing size conservation with increasing age for the experimental and control groups, respectively. Figure 1 permits the observation that in the experimental group size conservation is expressed by *Ss* as young as 6½ years of age and is evident at each successive age thereafter. On the other hand, Figure 2 indicates that the singular instance of size conservation in the control group does not appear until after 7 years of age. It should be acknowledged at this point that this control *S* did not evidence an adequate concept of "bigger" in the test following the conservation problem, a situation inconsistent with the hypothesis. Since, in response to the conservation inquiry, this *S* gave a compensatory explanation as well as a logical-necessity rationale, one is tempted to conjecture that this *S* actually had a multidimensional concept of "bigger" but simply made an error of volumetric judgment between the differently shaped test objects utilized in the test of that concept. Including this discrepant result, a comparison of the frequency of size-conserving responses given by all *Ss* of the experimental and control groups evidencing an adequate concept of "bigger" with those *Ss* failing to evidence this concept reveals a difference significant beyond the .01 level. Nevertheless, the above case, when taken at face value, stands as a rather weighty exception to this statistically significant relationship.

DISCUSSION

It seems clear that the training did affect the difference between the experimental and control *Ss* in regard to the adequacy of their concept of "bigger" that was necessary to test the hypothesis. It appears, furthermore, that this training affected a significant increase in the incidence of size conservation in response to the question "Is one bigger than the other?" However, this training seemed to facilitate the expression of

conservation for the older group only. This finding is, of course, consistent with our position that an adequate concept of "bigger" is a necessary but not sufficient condition for the manifestation of size conservation in this experimental condition.

Several investigators of size conservation have asked questions other than "Is one bigger than the other?" and, therefore, their data are not confounded with the adequacy of the concept of "bigger." Nevertheless, as Braine and Shanks (1965) have pointed out, those questions are still ambiguous in that some *Ss* may approach the task with a phenomenological set while others may take a more objective set. The fact that Braine and Shanks deliberately induce an objective set in their *Ss* accounts, in part, for their *Ss* expressing size conservation at an earlier age than the *Ss* of those investigators who have not addressed themselves to this matter. (Gruen, 1966, has emphasized that an additional reason for this discrepancy in age of attaining conservation is that Braine and Shanks do not require logical explanations in addition to correct responses to the initial conservation question as do the Piagetian investigators.) In addition to the point made by Braine and Shanks, our data emphasize that in order to reduce the variance among the data of future investigations of size conservation and thus permit more meaningful comparison with the other instances of conservation of amount as well as conservation of weight and volume, it is mandatory that the size conservation question be completely free of propaedeutic concepts not yet attained by the youngest of the *Ss* to whom the question is addressed.

REFERENCES

Braine, M. D. S., and Shanks, B. L. The development of conservation of size. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 227–242.

Gruen, G. E. Note on conservation: Methodological and definitional considerations. *Child Development,* 1966, **37**, 977–983.

Hayes, W. L. *Statistics for psychologists.* New York: Holt, 1963.

Lumsden, E. A., and Poteat, B. W. S. The salience of the vertical dimension in the concept of "bigger" in five- and six-year olds. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 404–408.

Smedlund, J. The acquisition of conservation of substance and weight in children: II. External reinforcement of conservation of weight and of the operations of addition and subtraction. *Scandinavian Journal of Psychology,* 1961, **2**, 71–84.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 92–105 (1969)

Partial Reward and Double Alternation Learning in Children[1]

PETER WATSON
Trent University

The effects of partial reward on double alternation (DA) learning in Grade II children were investigated in two experiments. In the first, *Ss* attempted to learn the DA problem under conditions of 50 and 100% social reward for correct responses. Although reward schedule did not affect the rate of learning, it did affect the number of *Ss* who learned; significantly more *Ss* solved the problem under 50% reward than under 100% reward. In the second experiment, all *Ss* attempted to learn the DA problem under 50% reward conditions after either success or failure pretraining. Pretraining did not affect the number of *Ss* who learned; however, the success-pretraining *Ss* who learned did so significantly more rapidly than those *Ss* who learned following failure pretraining. The results were interpreted in terms of frustration resulting from nonattainment of expected social reward; efficiency in solving the DA problem was seen to be a joint function of individual differences in reaction to frustration and the particular structure of the DA problem.

A number of recent studies with young children (e.g., Bruning, 1964; Pederson, 1967; Ryan and Watson, 1966; Watson, 1968) have demonstrated that partial reinforcement (e.g., 50%) for a simple instrumental motor response leads to faster performance than is the case when continuous reinforcement (100%) is given. Ryan and Watson (1968) have attributed these findings to frustration resulting from nonattainment of an expected reward. According to this frustrative-nonreward hypothesis (Amsel, 1958; see also Spence, 1960, Chap. 6), nonattainment of an expected reward leads to frustration, an aversive emotional condition which contributes to drive level and also possesses cue properties. The increment in motivation in *Ss* experiencing nonreward enhances their performance relative to that of *Ss* who never experience nonreward; i.e., those performing under 100% reward conditions, provided that the correct instrumental response is dominant.

While the bulk of evidence obtained thus far indicates that the frus-

[1] This work was supported in part by National Research Council of Canada Grant No. APA-250. The author is grateful to the Board of Education of the City of Peterborough for its co-operation, and to Mrs. Teresa Kinnear and Mr. Martyn Thomas, who served as *Es* in Exps. I and II, respectively. Thanks are also due to Dr. T. J. Ryan for his criticisms of a prepublication draft.

trative-nonreward hypothesis can account for children's behavior in very simple performance situations involving nonreward, little research has been directed toward the question of its applicability to more complex situation. Two studies (Lobb, 1966; Riese and Lobb, 1967) have found that discrimination learning is facilitated by pretraining in which extra nonreinforcements are given to the negative stimulus; this facilitative effect is greater than that resulting from pretraining in which extra rewards are given to the positive stimulus. In terms of the frustrative-nonreward hypothesis, nonreward evokes frustration which in pretraining becomes conditioned to the negative stimulus, producing an avoidance tendency which generalizes to the subsequent discrimination learning situation.

This brief and selective survey of the literature pertaining to frustrative nonreward in children suggests, then, that frustration has both energizing and directive functions. The studies reviewed also contain the suggestion that as the complexity of the task increases, the relative importance of the energizing and the directive functions may change, so that the cue properties of frustration may become somewhat more important as the task becomes more difficult or complex. The aim of the present study was to investigate the effects of partial reward for correct responses on children's double alternation (DA) learning, a relatively complex form of learning. DA learning, since its initial use by Hunter (1920), has been recognized as a prototype of processes often designated as symbolic behavior, cognitive functioning, or reasoning. According to Hodges (1956, p. 59), the subject "must respond in terms of temporal relations under conditions where no differential sensory cues exist." Several investigators (Gellerman, 1931; Hodges, 1954, 1956; Hunter and Bartlett, 1948) have reported moderately high correlations between intelligence or mental age or both and the ability to solve the DA problem. If, as seems likely, DA learning involves a process or processes which may properly be described as cognitive, the use of DA learning may prove valuable in investigating the effects of frustrative nonreward on children's complex behavior.

EXPERIMENT I

The specific aim of Exp. I was to determine the effects of reward schedule (50 vs. 100%) on DA learning in Grade II children. One major problem in a study of this type (see Osler and Shapiro, 1964) is that of ensuring that schedule-of-reward effects are not confounded with information feedback; i.e., that groups receiving different percentages of reward receive equal amounts of information concerning the correctness or incorrectness of their responses. One group of *Ss* (Group 100) received continuous social reward for correct responses in the DA situation, while another group (Group 50) was rewarded for a random half of their cor-

rect responses; knowledge of results was provided on all trials for all *Ss*. It was expected, on the basis of the frustrative-nonreward hypothesis, that the use of random partial reward in Group 50 would produce a somewhat higher motivational level than was the case for Group 100; in addition, Group 50 would be performing in the presence of whatever directive cues were associated with the nonreward-produced frustration. The unknown nature and effects of these cues precluded any prediction regarding them, but necessitated examination of the response-by-response behavior of the *Ss*, since their effects would presumably depend on the *Ss*' position in the DA sequence. The expected effects of increased motivation on Group 50 relative to Group 100 were also unclear. On one hand, increases in motivation might be expected to facilitate performance. On the other hand, assuming the validity of an inverted-U-shaped relationship between arousal level and performance (Bindra, 1959; Malmo, 1959; Spence, 1956) too great an increment in motivation might be expected to interfere with performance on this complex task (see Yerkes and Dodson, 1908). Because of these considerations, no specific predictions were made concerning the outcome.

Method

Subjects. Fifteen male and twenty-five female Grade II children served as *Ss*. They were assigned randomly to two groups: Group 50, composed of 7 males and 13 females (mean age = 92.2 months), and Group 100, composed of 8 males and 12 females (mean age = 89.2 months).

Materials and Procedure. The experimental task was a modification of the Double Alternation Card Test as used by Hodges (1954, 1956). The *Ss* were brought individually to the testing room and seated facing the *E* with a table between them. The *E* then dealt two ordinary playing cards, one red and one black, face up on the table, and said:

> "See, this card is red, and this one is black. Now, I'll turn them over, and you pick out the red one. Just pick out one of the cards."

If *S* failed to pick the red card, he was discarded and another *S* was obtained. If he was successful, *E* proceeded with the experiment, giving the following instructions:

> "Now, from now on, I won't show you which card is red. You'll have to guess which one it is. There is a way to be right all the time; let's see if you can find the way to be right every time. Remember, just pick one card, and see whether you're right or wrong."

Cards were then dealt out face down two at a time, one red and one black. The double alternation sequence was such that the red card was on the right twice and then on the left twice (RRLL). On each presentation of a pair of cards, *S* responded by choosing one. Each trial consisted of one RRLL sequence (four responses), and the procedure was continued either until 25 trials (100 responses) had been given or until *S* obtained two successive correct trials. The interresponse interval was about 8 seconds. New cards were employed after every two trials, in order to minimize the possibility that *S*s would learn to identify the correct card on the basis of accidental marks on the backs of the cards.

*S*s in Group 100 were socially rewarded ("Good") immediately after each correct response, while *S*s in Group 50 were socially rewarded on half of their correct responses. Arrangement of rewards and nonrewards for Group 50 was according to a pre-arranged random schedule. The *E* was silent after all incorrect responses for both groups. Since *S*'s response indicated to him whether he was right or wrong, the groups were equated in terms of knowledge of results. After each *S* had either achieved the criterion for learning or received 25 trials, he was questioned to determine whether he could verbalize the principle of solution, told he had done very well, and taken back to the classroom. A record of each response—whether it was correct, and if correct, whether it was rewarded—was kept for each *S*.

Results

Table 1 presents the number of *S*s who learned within 25 trials and the number who failed to learn, in Group 50 and Group 100, as well as the mean number of trials to criterion for each group. It will be noted

TABLE 1

NUMBER OF *S*S WHO LEARNED AND FAILED TO LEARN, AND MEANS AND STANDARD DEVIATIONS OF TRIALS TO CRITERION FOR GROUPS 50 AND 100

Group	Learners	Nonlearners	Trials to criterion[a]	
			$\bar{X}$	*SD*
50	14	6	13.6	9.5
100	7	13	19.4	8.4

[a] Assuming 25 trials to criterion for nonlearners.

that 14 *S*s in Group 50 learned, compared with half that number for Group 100 ($X^2 = 4.91$; $df = 1$; $p < .05$). Thus, it appears that giving random partial reward for correct responses facilitated DA learning.

relative to the case where all correct responses were rewarded. This difference is also reflected in the mean trials to criterion for each group as a whole, where the Group 50 mean was significantly lower than that for Group 100 ($F = 4.81$; $df = 1,38$, $p < .05$). These results do not, however, imply that *Ss* who solved the DA problem under 50% reward did so more quickly than those who learned under 100% reward; when learners only were considered, analysis of trials to criterion indicated no significant difference ($F = .01$; $df = 1,19$) between Group 50 ($n = 14$; $\bar{X} = 8.8$; $SD = 6.8$) and Group 100 ($n = 7$; $\bar{X} = 9.2$; $SD = 5.6$).

In order to determine why 50% reward facilitates DA learning, the data of Group 50[2] were subjected to a response-by-response analysis. Under the 50% reward condition, three outcomes of a response were possible: The response could be incorrect (I); it could be correct and rewarded (C-R); or it could be correct but not rewarded (C-NR). Since each of these outcomes could affect subsequent responses, their effects should be reflected in tendencies to "stay" or "shift," depending on whether *S* tends to repeat his previous response or not. Most previous investigations which have employed this line of reasoning (e.g., Rieber, 1966) have considered only the first response following a particular outcome. Because of the repetitive nature of the DA problem sequence, it is essential, in a response-by-response analysis, to consider *S*'s position in the sequence. The use of groups of three successive responses (LRR, LLR, RLL, or RRL), rather than two successive responses as used by Rieber (1966), is necessary and sufficient to establish *S*'s position at any point in the DA sequence. In the present study, both the first and second responses following each type of outcome were examined. For example, a response sequence "right" (C-NR), "left," "right," would be classified as a "shift" on the first response following the correct nonrewarded response, and a "stay" on the second response following. All responses except the last two prior to attaining criterion (or, in the case of nonlearners, responses 99 and 100) were tabulated according to this system. Following this, "shift ratios" were calculated for each type of outcome for each *S*; each shift ratio is the proportion of "shift" responses to the total number ("shift" plus "stay") of responses for each type of outcome. *Ss* were then divided into subgroups of fast learners (FL: $n = 8$), slow learners (SL: $n = 5$), and nonlearners (NL: $n = 6$); an arbitrary cut-off point of 10 trials to criterion was used to differentiate the FL and SL subgroups. After testing to ensure that variances were homogeneous within each type of response outcome, simple-randomized analyses of variance were used to test the significance of the differences among the

[2] With the exception of one *S* who solved the problem in one trial. This *S* was excluded because of insufficient data for use in a response-by-response analysis.

FL, SL, and NL subgroups for each response outcome. An insufficient number of *Ss* who learned in Group 100 prevented a similar breakdown and analysis of the response data for that group. Table 2 presents the shift ratio data for each subgroup of Group 50, as a function of type of response outcome, and the results of the analyses of variance.

The results of the analyses indicate that the three subgroups differ significantly only with regard to the second response following C-NR, a correct but nonrewarded response. *Ss* who solve the DA problem in 10 trials or less show a strong tendency to shift two responses after being nonrewarded ($\bar{X} = .79$), while *Ss* who solve the problem slowly or not at all show little tendency in either direction.

TABLE 2
SHIFT RATIO DATA FOR FL, SL, AND NL SUBGROUPS OF GROUP 50 AS A FUNCTION OF RESPONSE OUTCOME

		First response following:						Second response following:					
		C-R		C-NR		I		C-R		C-NR		I	
Subgroup	*n*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
FL	8	.53	.35	.59	.28	.66	.37	.64	.30	.79	.40	.53	.33
SL	5	.58	.40	.57	.35	.60	.32	.45	.25	.42	.25	.47	.26
NL	6	.53	.25	.51	.26	.67	.36	.57	.28	.53	.28	.48	.25
F-ratio		.17		.48		.80		2.61		7.66*		.30	

* $df = 2,16$; $p < .05$.

The final question of interest with regard to the results of Exp. I relates to the ability to verbalize the solution of the DA problem. Not surprisingly, no correct verbalizations occurred among the nonlearners of either Group 50 or Group 100. Among the learners, verbalization occurred in less than 30% of the *Ss*: in Group 50, four of 14 learners were able to verbalize the correct solution, while two of seven learners in Group 100 stated the solution correctly. It is obvious that reward schedule did not affect ability to verbalize the correct solution of the DA problem once it had been solved.

Discussion

The results of Exp. I indicate that DA problem solution in Grade II children is facilitated by giving partial reward for correct responses; schedule of reward had no effect on the ability to verbalize the basis of the solution. Analysis of the response-by-response data suggests that such facilitation may be related to *Ss'* reactions to the various response

outcomes involved in the partial-reward situation. Of the three types of response outcomes, both correct nonrewarded (C-NR) responses and incorrect (I) responses may be conceived of as eliciting frustration. In the first case, *S* has learned to expect reward for making the correct response, and frustration occurs when reward is withheld. In the second case, *S* in making a response expects that it will be correct[3] and rewarded; an incorrect response disconfirms that expectancy. Some theorists (e.g., Nuttin, 1968) argue strongly for a clear conceptual distinction between reward-nonreward and success-failure. The probable involvement of disconfirmation of an expectancy of reward in both cases outlined here implies that a complete distinction may be at least partially invalid. It is possible, of course, that a distinction in terms of the strength of frustration resulting from the two operations may be necessary; it could be assumed for example, that failure-produced frustration is stronger than frustration resulting from nonreward.

To the extent that a tendency to shift reflects the operation of a frustration-produced avoidance reaction, the relatively high shift ratios of the FL, SL, and NL subgroups of Group 50 on the first response following an incorrect response—shift ratios ranging from .60 to .67—suggest that all types of *Ss* experienced an incorrect outcome—"failure" in Nuttin's (1968) terms—as frustrating, and that this type of frustration was elicited quite rapidly. The relatively lower shift ratios two responses after an incorrect response—ranging from .47 to .53—indicated an equally rapid dissipation of this type of frustration for all three subgroups.

For correct nonrewarded responses, the other source of frustration in the partial-reward situation, only the FL subgroup demonstrated any avoidance tendency; this was apparent primarily on the second response following the correct nonrewarded response. It may be postulated from these observations that, while frustration resulting from an incorrect response evokes a short-lived avoidance tendency in most *Ss*, a relatively longer-lasting avoidance tendency is evoked in some *Ss* by nonattainment of an expected social reward, and not at all in others. That reactivity to nonreward is characteristic of *Ss* who solve the DA problem quickly is clear; less obvious is the nature of the relationship between problem-solving efficiency and susceptibility to frustrative nonreward.

Two possibilities seem open. The first is that fast learners are distinguishable from slow learners and nonlearners by an intellectual factor; they may solve the DA problem faster because of their greater capacity for cognitive manipulation, greater memory span, etc. Because these *Ss*

[3] Assuming of course that any given response does not serve solely information-gathering purposes.

are more capable, the argument would run, they have a higher expectancy for success, and for praise, than is the case for other *Ss*, and thus nonattainment of the expected social reward should be relatively more frustrating and should produce a stronger avoidance tendency. This argument in essence treats the strong shift tendency of the FL subgroup as a more or less irrelevant by-product of their problem-solving skill.

Two facts tend to disconfirm this hypothesis. First, attribution of ability to solve the DA problem rapidly to inherent intellectual superiority, besides begging the question, implies that partial reward does not affect problem-solving efficiency, and that both Group 100 and Group 50 should contain roughly the same proportion of learners and nonlearners. Second, the argument implies that fast learners should demonstrate greater avoidance tendencies than other *Ss* following incorrect as well as correct but nonrewarded responses. The data fail to support either implication.

The second possibility is the obverse of the first; i.e., rather than skill in problem-solving producing susceptibility to frustrative nonreward, susceptibility to frustrative nonreward may produce an apparent problem-solving skill. To clarify this point, it is necessary to examine the structure of the DA problem. In an extended DA sequence (e.g., RRLLRRLL), a variety of rules or strategies can be devised which will predict the correct response (R or L) on any given presentation of the two stimuli. These strategies employ the correctness or incorrectness and position of the two responses preceding the response to be predicted. It can be shown, however, that the most simple strategy, which subsumes all others, consists of two rules: (a) If the *n*th response was correct, "shift" on the *n*th + 2 response; (b) if the *n*th response was incorrect, "stay" on the *n*th + 2 response. It should be noted that this is a logical as opposed to psychological strategy, involving only the question of repeating or changing the *second* response after a given response outcome; according to the logic of this strategy, the response immediately following a given outcome is irrelevant if these rules are employed. It follows from these considerations that DA learning will be facilitated by anything which arouses tendencies either to shift two responses after being correct or to stay two responses after being incorrect. In the present experiment, it appears possible that the elicitation of a nonreward-produced avoidance tendency in some *Ss* may have essentially forced these *Ss* into a DA pattern of responding—a conclusion supported to some extent by the failure of partial reward to produce increased ability to verbalize the correct solution among those *Ss* who learned. Other *Ss*, who were not frustrated by nonattainment of the expected reward, would then solve the DA problem relatively slowly or not at all.

EXPERIMENT II

The results of Exp. I indicate that partial reward for correct responses facilitates DA learning in some children, and suggest that the facilitative effect is due to a relatively long-lasting avoidance response produced by nonreward. Experiment II was designed to manipulate the strength of frustration resulting from nonreward and failure, by means of success and failure pretraining. It was assumed (see Cromwell, 1963) that prior experience of success in a situation similar to the DA situation would increase *S*s' expectancy of success and social reward in the subsequent DA learning task. Nonreward would then elicit stronger frustration, a greater avoidance tendency two responses after a correct nonrewarded response, and relatively high efficiency in solving the DA problem. Conversely, prior failure experience should result in a lower expectancy of success and reward, weaker frustration resulting from nonreward, and relatively low problem-solving efficiency.

Method

Subjects. *S*s were 25 male and 15 female Grade II students assigned randomly to two experimental groups. The "success" group (Group S) consisted of 12 males and 8 females (mean age = 93.0 months), while the "failure" group (Group F) consisted of 13 males and 7 females (mean age = 91.8 months).

Materials and Procedure. With the exception of the pretraining phase of the experiment, the procedure was identical to that employed with Group 50 of Exp. I. The pretraining phase was conducted as follows. After giving the instructions in which *S* was informed that his task was to pick out the red card, *E* gave two trials (eight responses) using cards of only one colour. For Group S, red cards were employed while Group F were dealt only black cards. Thus, Group S made eight "correct" responses, while Group F made eight "incorrect" responses. All Group S responses were socially rewarded, while Group F responses were never rewarded. Following completion of the eight pretraining responses, *E* said:

> "Now we'll try a different one. Remember, there's a way to find the red card every time."

E then proceeded to administer DA trials until a criterion of two consecutive correct trials had been achieved or until 25 trials had been given. For all *S*s in both groups, correct responses were socially rewarded on a

TABLE 3
NUMBER OF FL, SL AND NL *Ss* IN GROUPS S AND F

	FL	SL	(Total L)	NL
Group S	9	1	(10)	10
Group F	3	5	(8)	12
Total	12	6	(18)	22

prearranged random 50% schedule. No attempt was made to determine whether *Ss* could verbalize the correct solution to the DA problem.

Results and Discussion

The *Ss* of Exp. II were classified as fast learners (FL), slow learners (SL), or nonlearners (NL) according to the same criteria used in Exp. I. The number of *Ss* in the FL, SL, and NL subgroups of Group S and Group F are presented in Table 3. The first test was performed on the number of *Ss* who learned (FL plus SL) and who failed to learn (NL) in the two groups; no significant difference between Group S and Group F was found ($X^2 = .40$; $df = 1$). The groups appeared to differ, however, with regard to the relative sizes of the FL and SL subgroups. Because of the small cell frequencies involved in this comparison, the Fisher exact probability test (Ferguson, 1959, pp. 173–174) was used to assess the difference. An obtained exact probability of .0316 indicated that the groups differed with regard to their relative proportions of fast and slow learners. This difference is also reflected in the mean trials to criterion of both the total groups and of just the learners within each group (Table 4). In both cases, the mean trials to criterion is lower for Group S than

TABLE 4
MEAN NUMBER OF TRIALS TO CRITERION FOR GROUPS S AND F

	Learners only			All subjects		
	n	$\bar{X}$	*SD*	*n*	$\bar{X}$	*SD*
Group S	10	6.4	4.0	20	15.7	10.0
Group F	8	12.5	5.4	20	20.0	10.0

for Group F; the difference is significant only when just those who learned are considered ($F = 5.30$; $df = 1,16$; $p < .05$).

These results suggest that while pretraining does not affect the number of *Ss* who learn or fail to learn, it does affect the speed of learning. *Ss* who solve the DA problem after success pretraining do so more rapidly than those who solve it after failure pretraining. If the assumption that

the type of pretraining experience influences *Ss*' expectancy of reward is valid, it follows from the frustrative-nonreward hypothesis that *Ss* who are given success pretraining should subsequently experience stronger frustration in the partially rewarded DA learning situation than is the case for *Ss* given failure pretraining. Thus, it appears that speed of problem solution is a function of the strength of the nonreward-produced frustration involved in the problem situation.

Shift ratios for each type of response outcome were calculated for each *S* following the method specified in Exp. I. Because of the small number of *Ss* in some of the FL and SL categories of Groups S and F, the two groups were combined in order to analyze for differences among FL, SL, and NL subgroups. The mean shift ratios of the three subgroups are presented in Table 5.

TABLE 5
SHIFT RATIO DATA FOR FL, SL, AND NL SUBGROUPS OF GROUPS S AND F COMBINED, AS A FUNCTION OF RESPONSE OUTCOME

		First response following:						Second response following:					
		C-R		C-NR		I		C-R		C-NR		I	
Subgroup	*n*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
FL	12	.70	.37	.64	.36	.72	.35	.50	.35	.69	.32	.43	.20
SL	6	.45	.28	.47	.24	.71	.36	.60	.30	.57	.32	.51	.25
NL	22	.60	.20	.56	.24	.69	.25	.48	.14	.49	.17	.46	.17

Because of marked heterogeneity of variance in the shift ratio data on the first and second response following C-R and on the second response following C-NR, differences among the FL, SL, and NL subgroups on these three types of response outcomes were assessed by means of median tests (Ferguson, 1959, p. 267). The remaining differences were assessed by means of simple-randomized analyses of variance. Table 6 presents the results of the tests of homogeneity, along with the results of the appropriate tests of significance among the FL, SL, and NL subgroups.

Of the six tests, only that for the second response following a correct nonrewarded response was significant. The FL subgroup mean was relatively high, while those of the SL and NL subgroups were lower. The general similarity of these results to the corresponding results of Exp. I should be noted, particularly the high shift ratios of the FL subgroup for the second response following a nonrewarded correct response, and the high shift ratios of all subgroups immediately following an incorrect response.

TABLE 6

RESULTS OF F_{max} TESTS, ANALYSES OF VARIANCE, AND MEDIAN TESTS FOR FL, SL, AND NL SUBGROUPS OF GROUPS S AND F COMBINED, AS A FUNCTION OF RESPONSE OUTCOME

	First response following:			Second response following:		
	C-R	C-NR	I	C-R	C-NR	I
F_{max}	3.50	2.16	2.16	6.00	3.33	2.00
df	11,21	11,21	4,21	11,21	11,21	4,21
p	.01	n.s.	n.s.	.01	.01	n.s.
Analysis of variance (*F*)	—	1.20	.25	—	—	.50
Median test (X^2)	3.81	—	—	2.30	6.21[a]	—

[a] $df = 2$; $p < .05$.

GENERAL DISCUSSION

The results of the two experiments indicate that while schedule of reward affects the number of *S*s who solve the DA problem but not the speed with which it is solved, pretraining designed to manipulate *S*s' expectancy of success and failure affects the speed with which it is solved but not the number of *S*s who solve it. Thus, in terms of the frustrative-nonreward hypothesis (Amsel, 1958), it may be that for some *S*s the mere presence or absence of frustrative nonreward determines whether or not the solution will occur, while the strength of nonreward-produced frustration determines the rate of solution. This latter conclusion is further illustrated when the results of Exp. II are compared with those of Group 50 in Exp. I. It may be seen, first, that success pretraining did not increase the number of *S*s who solved the problem relative to Group 50 in Exp. I, which was essentially a no-pretraining group. The mean trials to criterion for Group S of the present experiment ($\bar{X} = 6.4$) was, however, somewhat lower than that for Group 50 of the previous experiment ($\bar{X} = 8.8$). Because of the dangers inherent in interexperimental comparisons, more research is needed before placing too much emphasis on this point.

The data also indicate that solution of the DA problem is related to the specific action of nonreward two responses after the frustrating event. A conclusion that partial reward will facilitate any type of complex pattern learning seems unwarranted at the present time, because the present results rest on a particular combination of individual reactions to nonreward and the specific structure of the DA problem. Different problems—for example, triple alternation—might give entirely different results. Since in the case of triple alternation the optimum response strategy is to "shift" three responses after a correct response and to "stay" three

responses after an incorrect response, it might be expected that partial reward would facilitate triple alternation learning only for *Ss* who exhibit an extremely long-lasting avoidance response to frustration; for *Ss* who show a brief response to frustration, partial reward should have little effect on their ability to solve the triple alternation problem.

The present studies were conceived and are interpreted within the framework of Amsel's (1958) hypothesis of frustrative nonreward. While alternative interpretations based on associative or motivational factors other than frustration might be advanced to account for the obtained results, the form such alternatives would take is not particularly clear. What is clear is that any interpretation of these results will have to take into account the crucial role of individual differences in reaction to nonreward. The current status of most hypotheses of psychological functioning, including the one employed here, is such that individual differences are usually regarded as a source of error to be minimized or eliminated. The present results suggest the need not necessarily to eliminate but to predict and employ individual differences in the design of learning tasks.

The above statements imply that it may be possible to predict whether or not partial reward will facilitate learning if the individual's particular response to frustration produced by nonreward or failure is known, and if the problem is or can be structured to take advantage of this response. Presumably, the question of an individual's particular response to frustration can be settled by means of techniques such as the ones used in this study, or perhaps by more direct measures including physiological ones. The task of creating an optimum structure for a problem which must be learned is probably more difficult, especially as the type of problem increases in complexity. For example, a problem with a relatively simple structure, such as the DA problem (RRLL), would be much easier to make predictions for than would a similar problem with a more complex inherent structure (e.g., RLLRRL). Finally, it may be extremely difficult to impose a structure on a very large number of learning problems which an individual encounters outside the laboratory; although the difficulties may be enormous, the returns from such an effort may ultimately prove of great value for increasing learning efficiency.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

BINDRA, D. *Motivation: A systematic reinterpretation*. New York: Ronald Press, 1959.

BRUNING, J. L. The effects of magnitude of reward and percentage of reinforcement on a lever-movement response. *Child Development*, 1964, **35**, 281–285.

CROMWELL, R. L. A social learning approach to mental retardation. In N. R. Ellis (Ed.) *Handbook of Mental Retardation.* New York: McGraw-Hill, 1963.

FERGUSON, G. A. *Statistical Analysis in Psychology and Education.* New York: McGraw-Hill, 1959.

GELLERMAN, L. W. The double alternation problem: II. The behavior of children and human adults in a double alternation temporal maze. *Journal of Genetic Psychology,* 1931, **39,** 197–226.

HODGES, A. A developmental study of symbolic behavior. *Child Development,* 1954, **25,** 277–280.

HODGES, A. Double alternation: a measure of intelligence. *Journal of Consulting Psychology,* 1956, **20,** 59–62.

HUNTER, W. S. The temporal maze and kinesthetic sensory processes in the white rat. *Psychobiology,* 1920, **2,** 1–18.

HUNTER, W. S., AND BARTLETT, S. C. Double alternation behavior in young children. *Journal of Experimental Psychology,* 1948, **38,** 558–567.

LOBB, H. O. Visual discrimination learning in imbecile children with nonreinforcement of irrelevant tendencies. *American Journal of Mental Deficiency,* 1966, **70,** 753–762.

MALMO, R. B. Activation: A neurophysiological dimension. *Psychological Review,* 1959, **66,** 367–386.

NUTTIN, J. *Reward and punishment in human learning.* New York: Academic Press, 1968.

OSLER, SONIA, F., AND SHAPIRO, SANDRA L. Studies in concept attainment: IV. The role of partial reinforcement as a function of age and intelligence. *Child Development,* 1964, **35,** 623–633.

PEDERSON, D. R. Associative vs. motivational interpretations of reward percentage effects on children's performance. *Psychonomic Science,* 1967, **8,** 139–140.

RIEBER, M. Response alternation in children under different schedules of reinforcement. *Psychonomic Science,* 1966, **4,** 149–150.

RIESE, R., AND LOBB, H. O. Discrimination learning in retarded children: Nonreward vs. reward. *American Journal of Mental Deficiency,* 1967, **71,** 536–541.

RYAN, T. J. Instrumental performance as related to several reward schedules and age. *Journal of Experimental Child Psychology,* 1966, **3,** 398–404.

RYAN, T. J., AND WATSON, P. Children's response speeds as a function of sex and verbal reinforcement schedule. *Psychonomic Science,* 1966, **6,** 271–272.

RYAN, T. J., AND WATSON, P. Frustrative nonreward theory applied to children's behavior. *Psychological Bulletin,* 1968, **69,** 111–125.

SPENCE, K. W. *Behavior Theory and Conditioning.* New Haven: Yale University Press, 1956.

SPENCE, K. W. *Behavior Theory and Learning: Selected Papers.* Englewood Cliffs: Prentice-Hall, 1960.

WATSON, P. Response measure effects in children as a function of type and schedule of reward. *Psychonomic Science,* 1968, **13,** 217–218.

YERKES, R. M., AND DODSON, J. D. The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative and Neurological Psychology,* 1908, **18,** 459–482.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 106–117 (1969)

The Effects of Differential Amounts of Stimulus Familiarization on Choice Reaction Time Performance in Children[1]

FRANK D. MILLER
University of South Dakota

Groups of first-grade children were exposed during a familiarization phase to 0, 1, 10, 20, or 40 3-second presentations of a red or a green light. On 49 subsequent choice reaction time trials, the familiarized stimulus (FS) and nonfamiliarized stimulus (NS) each served as the signal to respond on half of the trials and were presented according to one of two sequences. Both sequences involved 50% alternation and 50% repetition of the two types of stimuli. Release but not travel speeds were sensitive to the effects of familiarization. Release speeds were faster to the NS than to the FS for *Ss* given 20 or 40 familiarization exposures but not for those given 0, 1, or 10 exposures. The magnitude of this effect was comparable for *Ss* given 20 or 40 exposures. A change effect (CE) was found for both release and travel speeds except in the case of travel speed for Group 0. The CE findings can be accounted for by either a stimulus interpretation or a modified response interpretation.

Several investigators have employed a similar two-phase paradigm to examine the effects of familiarization on response speeds in children. During a familiarization phase, *S* attends without responding to repeated presentations of a stimulus. This phase is followed by a motor task in which *S* responds to the onset of either the familiarized stimulus (FS) or a nonfamiliarized stimulus (NS) not presented in familiarization.

A stimulus familiarization effect (SFE), indicated by faster response speeds to the NS than to the FS, has been obtained in all studies for the initiation segment of the response (Bogartz and Witte, 1966, Exp. II; Cantor and Cantor, 1964; Cantor and Fenson, 1968; Witte, 1965, 1967; Witte and Cantor, 1967) and also in those studies utilizing a composite

[1] This study is based on a dissertation submitted in 1966 to the Graduate College of the University of Iowa in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Institute of Child Behavior and Development. The writer wishes to express his sincere appreciation to Gordon N. Cantor and William J. Meyers for their thoughtful direction throughout the course of the investigation. Thanks are also due Gordon N. Cantor for his helpful suggestions regarding the manuscript.

response measure (Bogartz and Witte, 1966, Exp. I; Cantor and Cantor, 1965, 1966; Miller and Moffat, 1968). The SFE has not been obtained for the execution segment of the response except in one study by Witte and Cantor (1967) in which *Ss* responded to stimulus offset rather than stimulus onset.

Two studies have been concerned with the effect of amount of familiarization on the magnitude of the SFE. Cantor and Cantor (1966) employed a composite response measure and found that the magnitude of the SFE was comparable for groups given either 5 or 35 3-second familiarization exposures. In contrast, Cantor and Fenson (1968) obtained an initiation speed SFE for a group of *Ss* given 18 3-second familiarization exposures but not for groups of *Ss* given 5, 10, 12, 14, or 16 such exposures. This effect was demonstrated only in the first half of the motor task trials. In addition, no SFE was found for the execution speed measure. One purpose of the present study was to reexamine the effect of amount of familiarization on magnitude of the SFE by comparing groups given 0, 1, 10, 20, or 40 3-second familiarization exposures.

Another finding of interest in the familiarization studies has been the change effect (CE). The CE is characterized by faster response speeds on trials involving a stimulus change from the previous trial (i.e., NS following FS or FS following NS) than on trials involving no stimulus change from the previous trial (i.e., NS following NS or FS following FS). The CE has been obtained in studies (Bogartz and Witte, 1966, Exp. II; Cantor and Cantor, 1965, 1966) which have involved two combined conditions. The first condition was that *Ss* were required to make a different response to FS onset than to NS onset and the second was that the stimulus sequences involved predominantly more change than nonchange trials. In the Cantor and Cantor studies (1965, 1966), the required response was a differential response and the trial sequences involved approximately 75% change and 25% nonchange trials. The response in the Bogartz and Witte study (1966, Exp. II) was divided into a nondifferential initiation and a differential execution portion. Half of the *Ss* were given stimulus sequences which involved 75% change and 25% nonchange trials, while the remaining *Ss* were given sequences which involved 25% change and 75% nonchange trials. A CE was not found for initiation speeds regardless of the stimulus sequences employed. However, a CE was shown for execution speeds with the 75% change–25% nonchange sequences, whereas a reverse CE (i.e., speeds faster on nonchange than on change trials) was obtained for execution speeds with the 25% change–75% nonchange sequences.

In contrast to the differential response findings, the CE has not been found in studies (Bogartz and Witte, 1966, Exp. I; Cantor and Cantor,

1964; Miller and Moffat, 1968; Witte, 1967; Witte and Cantor, 1967) where *S* makes the same response (lever pulling or simple reaction time) to both stimuli. This finding holds regardless of the proportion of change to nonchange trials used.

Cantor and Cantor (1966) suggest a stimulus interpretation of the CE, hypothesizing that the CE may be due to the "novelty" inherent in changing from one stimulus to another from trial to trial within the motor task. On the other hand, Bogartz and Witte (1966) suggest a response interpretation. They argue that a *predominantly alternating* stimulus sequence in conjunction with a *differential* response induces an alternation tendency in the children which results in increased speeds in making an alternation response (change trials) and decreased speeds in making a repetition response (nonchange trials).

A second purpose of the study was to determine whether a CE could be obtained for the nondifferential initiation and the differential execution segments of a response when the stimulus sequences involve 50% alternation and 50% repetition of the NS and FS.[2] Cantor and Cantor's (1966) stimulus interpretation would lead one to expect a CE for both release and travel speeds. However, Bogartz and Witte's (1966) response interpretation suggests that a release speed CE should not be found, since this is the nondifferential portion of the response. Nor should one expect to find a CE for travel speed with the 50% alternation–50% repetition sequences employed.

METHOD

Subjects[3]

The *Ss* were 160 first-grade children obtained from the Grant Wood, Cleveland, and Taylor Elementary Schools in Cedar Rapids, Iowa. Ages of the *Ss* ranged from 6 years, 7 months to 8 years, 2 months, with a mean age of 7 years, 2 months.

Apparatus[4]

The apparatus used has been described in detail in a previous study (Bogartz and Witte, 1966, Exp. II). A circular glass aperture, 2 inches in diameter, was located in the center of the vertical front surface of a

[2] Cantor and Fenson (1968) have collected data relevant to this point. These data will be presented in a manuscript now in preparation.

[3] The writer is indebted to Ira J. Semler, Director of Research, and other personnel of the Cedar Rapids, Iowa Community School District for their cooperation in providing subjects.

[4] Thanks are due Merle Miller for assistance in construction and maintenance of the apparatus.

box. Three white light bulbs were located within the box and behind the aperture. One of these allowed for the presentation of a white light and the other two in combination with colored filters allowed for the presentation of a red or green light.

Three microswitches, two in posterior positions and one in an anterior position, were located on a horizontal surface extending forward from the bottom of the box. The anterior switch was mounted on the midline equidistant (7 inches) from the posterior switches. Three removable circular response buttons, one inch in diameter, were available for mounting on the switches. One was painted black, another red, and the other green.

Procedure

Each *S* was individually accompanied by *E* from the classroom to a dimly illuminated experimental room where he was seated in front of the apparatus. Each *S* was then instructed to watch carefully while a light came on and went off several times.

One-fifth of the *S*s were randomly assigned to each of five groups (i.e., Groups 0, 1, 10, 20, and 40) which were given 0, 1, 10, 20, or 40 presentations of either the red or green light, respectively, during the familiarization phase. In addition, *S*s in Groups 0, 1, 10, 20, and 40 were given 40, 39, 30, 20, or 0 presentations of the white light, respectively, *prior to* the presentations of the familiarization stimulus. Hence, each *S* received a total of 40 3-second stimulus presentations during the familiarization phase. Stimulus duration was controlled by a Hunter Decade Interval Timer. For half the *S*s in Groups 1, 10, 20, and 40, the red light served as the familiarization stimulus; the remaining *S*s in each of these groups were familiarized on the green light. Although the *S*s in Group 0 were not given presentations of the red or green light, half of them were randomly assigned to a "red" subgroup and the remaining *S*s to a "green" subgroup. For all *S*s, *E* said 'ready' before each stimulus presentation. An interval of 2, 3, or 4 seconds, timed with a stopwatch, occurred between the ready signal and light onset. Each of the three intervals occurred once within each successive block of three stimulus presentations. A 3-second interval always occurred on Trial 40. This order was common for all *S*s. A 5-second intertrial interval, also timed with the stopwatch, separated light offset and the next ready signal.

A 2-minute rest period followed the familiarization phase. During this period, *E* determined *S*'s preferred hand by asking him to mount the response buttons on the switches with the hand that he used for writing. For half the *S*s, the red button was mounted on *S*'s right and the green button on his left; for the remaining *S*s, the reverse spatial arrangement

was used. The black button was always mounted on the anterior switch and served as a start position.

The *S* was now told to put his preferred hand on the black button and to hold it down. The *E* explained that he would say 'ready' before either the red or green light came on and that *S* should release the anterior button and press the posterior button corresponding in color to that of the light as quickly as possible. The *S* was also instructed that, after pressing either button, he should return his hand to the black button and hold it down. The *E* asked *S* to demonstrate what he would do when a red light came on and then also what he would do when a green light came on.

Forty-nine choice reaction time trials were then administered according to one of two sequences counterbalanced for order of the NS and FS. Half of the *S*s in each of the ten amount of familiarization-color counterbalancing subgroups were given one sequence, and the remaining *S*s in each of these subgroups were administered a converse sequence. The order of stimuli occurring in one sequence was NFFNNNFF for the first eight trials, with the same eight-trial sequence being repeated five more times for a total of 48 trials. An NS was then presented on the last trial. The order of stimuli occurring in the converse sequence was FNNFFFNN for each of the six successive blocks of eight trials, with an FS being presented on Trial 49.

The *E* said 'ready' before each light presentation, using a restricted random order of 2-, 3-, and 4-second intervals between the ready signal and light onset. For the first trial, a 3-second interval was used. For the remaining 48 trials, each of the three intervals occurred on four occasions for each of the four possible sequential stimulus events (i.e., NS–FS, FS–NS, NS–NS, FS–FS). This restricted random order was common to all *S*s. The *E* depressed a silent push button at the back of the box which simultaneously activated the appropriate light and a Hunter KlocKounter. When *S* released the black button, the KlocKounter was deactivated and a Standard Electric Timer was simultaneously activated. Depression of either the red or green button by *S* turned off the stimulus light and stopped the standard timer. If *S* made an incorrect response, this was indicated by a signal light on the back of the apparatus and the trial was immediately repeated.[5] After each trial, *E* recorded *S*'s response latencies (i.e., release time and travel time) to the nearest 0.01 second. Release time was the amount of time taken by the *S* to lift his finger off the start button following stimulus onset. Travel time was the time taken to press the posterior button corresponding in color to that of the stimulus light after releasing the start button. The interval between

[5] There was a total of 38 corrected trials and these were fairly evenly distributed across the amount of familiarization groups. This total includes 12 FS–FS, 12 FS–NS, 8 NS–NS, and 6 NS–FS trials.

stimulus offset and presentation of the next verbal ready signal was approximately 9 seconds. Before the next ready signal, *E* manipulated a rotary switch which controlled the color of the stimulus light. This was done prior to each trial whether necessary or not, to avoid giving *S* clues about the color of the next light.

RESULTS

A reciprocal transformation was used to convert release and travel time on trials 2–49 to speed measures. A .05 significance level was used for all statistical tests.

Analyses Concerned with the SFE

For each *S*, mean release and travel speeds were computed separately for the NS and FS over three blocks of eight trials each. These means were then employed as the data in two analyses of variance, one for release and the other for travel speeds. In each overall analysis, a double extension of a Lindquist (1953) Type VII design was employed, with amount of familiarization and sequence as the between-*S*s main effects and NS vs. FS (i.e., the SFE comparison), stimulus color, and trial blocks as the within-*S*s effects.

Release speed. The following four effects were statistically significant in the overall release speed analysis: (a) NS vs. FS ($F = 29.25$, $df = 1$, 140 $p < .001$); (b) color of the stimulus ($F = 15.58$, $df = 1$, 140, $p < .001$); (c) trial blocks ($F = 59.70$, $df = 2$, 280, $p < .001$); and (d) the interaction involving NS vs. FS $\times$ Amount of Familiarization ($F = 5.35$, $df = 4$, 140, $p < .001$).

The NS vs. FS $\times$ Amount of Familiarization interaction is illustrated in Figure 1 where it may be seen that, for Group 0, speeds to the arbitrarily designated NS and FS were almost identical. For Groups 1 and

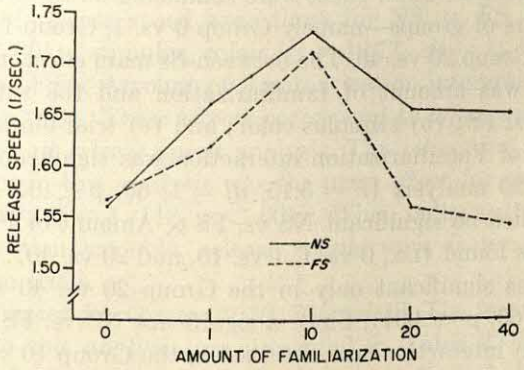


FIG. 1. Mean release speeds to the NS and FS for each amount of familiarization group.

10, speeds were slightly faster to the NS than to the FS; in Groups 20 and 40, speeds were markedly faster to the NS than to the FS. The obtained release speed differences (NS minus FS) for Groups 1, 10, 20, and 40 were .040, .020, .095, and .109, respectively. The significant main effect for NS vs. FS is reflected in Figure 2 in the generally faster speeds to the

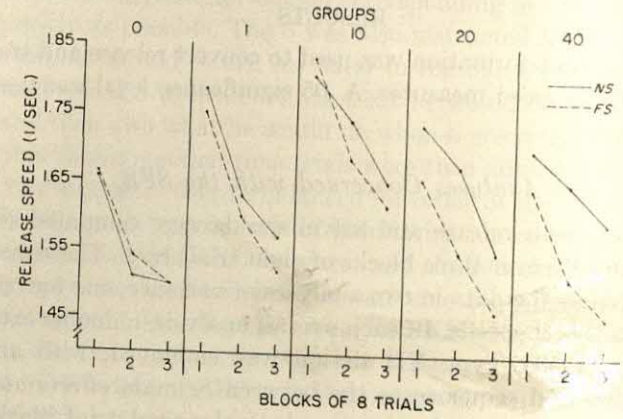


FIG. 2. Mean release speeds to the NS and FS as a function of trial blocks for Groups 0, 1, 10, 20, and 40.

NS than to the FS for Groups 1, 10, 20, and 40. The main effect for color of the stimulus resulted from the fact that the release speeds were faster on the average to the red than to the green light, regardless of which one served as the NS. The trial blocks main effect is attributable to the decrease in mean release speeds across trial blocks for all familiarization groups.

In order to follow up the NS vs. FS $\times$ Amount of Familiarization interaction, Lindquist (1953) Type VII analyses of variance extended to one additional between-*Ss* main effect were conducted on mean release speeds for specific pairs of groups—namely Group 0 vs. 1, Group 1 vs. 10, Group 10 vs. 20, and Group 20 vs. 40. The between-*Ss* main effect in each of these four analyses was amount of familiarization and the within-*Ss* effects were (a) NS vs. FS; (b) stimulus color; and (c) trial blocks. The NS vs. FS $\times$ Amount of Familiarization interaction was significant only for the Group 10 vs. 20 analysis ($F = 5.15$, $df = 1, 60$, $p < .05$). In the three analyses in which no significant NS vs. FS $\times$ Amount of Familiarization interaction was found (i.e., 0 vs. 1, 1 vs. 10, and 20 vs. 40), the NS vs. FS main effect was significant only in the Group 20 vs. 40 analysis ($F = 37.17$, $df = 1, 60$, $p < .001$). Since a significant NS vs. FS $\times$ Amount of Familiarization interaction was obtained in the Group 10 vs. 20 analysis, Lindquist (1953) Type VII analyses of variance extended to one ad-

ditional between-*Ss* variable were conducted on mean release speeds for Groups 10 and 20 separately. In each of these two analyses, the between-*Ss* main effect was sequence and the within-*Ss* effects were the same as those in the group comparison analyses. The *F* ratios for the NS vs. FS main effect were as follows: (a) Group 10—$F < 1.0$; and (b) Group 20—$F = 16.58$, $df = 1, 28$, $p < .001$.

These group comparison and follow-up analyses indicated that: (a) release speeds to the NS and FS were not significantly different for Groups 0, 1, and 10; (b) release speeds were significantly faster to the NS than to the FS for *Ss* given 20 and 40 familiarization trials; and (c) the magnitude of the release speed SFE was comparable for Groups 20 and 40.

Travel speed. The only statistically significant effect obtained in this overall analysis was the Amount of Familiarization × Trial Blocks interaction ($F = 2.49$, $df = 8, 280$, $p < .05$). No further consideration was given to this interaction, since it did not involve the NS vs. FS variable. No additional analyses concerned with the SFE were conducted on the travel speed data.

Analysis Concerned with the CE

Utilizing the data from trials 2–49 for each *S* in Groups 1, 10, 20, and 40, mean release and travel speeds were calculated for responses to an FS following an NS, to an NS following an FS, to an NS following an NS, and to an FS following an FS. These mean speeds were used as the basic data in two analyses of variance, one for each response measure. Both analyses involved a double extension of a Type VII design (Lindquist, 1953) in which amount of familiarization and stimulus sequence were the between-*Ss* variables and NS vs. FS, color of the stimulus, and change vs. no change from the previous trial were within-*Ss* variables.

Release speed for Groups 1, 10, 20, and 40. Table 1 shows the mean release speeds of interest. Three statistically significant effects in the analysis of release speed were those for NS vs. FS ($F = 35.53$, $df = 1, 112$, $p < .001$), stimulus color ($F = 12.77$, $df = 1, 112$, $p < .001$), and the NS vs. FS × Amount of Familiarization interaction ($F = 3.66$, $df = 3, 112$, $p < .05$). These effects correspond to what was shown previously in the overall release speed analysis. The remaining statistically significant effect in this analysis was the main effect for change vs. no change ($F = 26.20$, $df = 1, 112$, $p < .001$). When collapsed over the amount of familiarization variable, release speeds were faster on change than on nonchange trials.

Travel speed for Groups 1, 10, 20, and 40. The mean travel speeds that pertain to this analysis are presented in Table 2. The only significant statistic obtained was the *F* for the main effect of change vs. no change

TABLE 1
MEAN RELEASE SPEEDS FOR GROUPS 1, 10, 20, AND 40 TO THE FAMILIAR AND NONFAMILIAR STIMULUS ON CHANGE AND NONCHANGE TRIALS

Group	Sequential relation	Stimulus event		
		Familiar	Nonfamiliar	Mean
1	Change	1.660	1.689	1.675
	Nonchange	1.580	1.631	1.606
	Mean	1.620	1.660	
10	Change	1.735	1.739	1.737
	Nonchange	1.685	1.723	1.704
	Mean	1.710	1.730	
20	Change	1.585	1.674	1.630
	Nonchange	1.531	1.632	1.582
	Mean	1.558	1.653	
40	Change	1.577	1.664	1.621
	Nonchange	1.500	1.632	1.566
	Mean	1.539	1.648	

($F = 4.55$, $df = 1, 112$, $p < .05$). Regardless of amount of familiarization, travel speeds were faster on change than on nonchange trials.

Release and travel speeds for Group 0. For each S in Group 0, the data from trials 2–49 were used to compute mean release and travel speeds for responses to a red light following a green light, to a green light following a red light, to a red light following a red light, and to a green light following a green light. Since there were no familiarized and nonfamiliar-

TABLE 2
MEAN TRAVEL SPEEDS FOR GROUPS 1, 10, 20, AND 40 TO THE FAMILIAR AND NONFAMILIAR STIMULUS ON CHANGE AND NONCHANGE TRIALS

Group	Sequential relation	Stimulus event		
		Familiar	Nonfamiliar	Mean
1	Change	2.975	2.997	2.986
	Nonchange	2.934	2.925	2.930
	Mean	2.955	2.961	
10	Change	3.067	3.140	3.104
	Nonchange	3.001	3.057	3.029
	Mean	3.034	3.099	
20	Change	3.044	3.094	3.069
	Nonchange	3.032	3.009	3.021
	Mean	3.038	3.052	
40	Change	3.139	3.153	3.146
	Nonchange	3.135	3.114	3.125
	Mean	3.137	3.133	

ized stimuli for Group 0, color of the light was used as the basis for classifying change (i.e., R–G + G–R) and nonchange (i.e., R–R + G–G) trials. Two Lindquist (1953) Type VI analyses of variance, one for release and the other for travel speeds, were conducted on these data. The between-*Ss* variable was stimulus sequence and the within-*Ss* effects were color of the stimulus on the current trial and change vs. no change from the previous trial.

Table 3 contains the relevant means for both release and travel speeds. The only significant effect for release speeds was the main effect for change vs. no change ($F = 33.51$, $df = 1, 30$, $p < .001$). This effect is attributable to the occurrence of faster speeds on change than on nonchange trials. A significant main effect was found for stimulus sequence in the travel speed analysis ($F = 6.09$, $df = 1, 30$, $p < .05$). This effect has no direct bearing on the results of primary interest. None of the remaining effects was significant in the analysis of travel speed.

TABLE 3

MEAN RELEASE AND TRAVEL SPEEDS FOR GROUP 0 TO THE RED AND GREEN STIMULUS ON CHANGE AND NONCHANGE TRIALS

Response measure	Sequential relation	Stimulus event		
		Red	Green	Mean
Release speed	Change	1.627	1.597	1.612
	Nonchange	1.528	1.499	1.514
	Mean	1.578	1.548	
Travel speed	Change	3.302	3.334	3.318
	Nonchange	3.260	3.317	3.289
	Mean	3.281	3.326	

DISCUSSION

A finding of interest was the NS vs. FS × Amount of Familiarization interaction obtained in the overall analysis of release speeds. Significant SFEs were found for *Ss* given 20 or 40 3-second familiarization exposures but not for *Ss* given 0, 1, or 10 such exposures. These results suggest that a minimum of about 20 3-second familiarization exposures are required to produce the SFE. These findings are in agreement with those of Cantor and Fenson (1968) who obtained an initiation speed SFE for *Ss* given 18 3-second exposures but not for *Ss* given 16, 14, 12, 10, or 5 exposures. These results are in disagreement with those of Cantor and Cantor (1966) who found an SFE for *Ss* given 5 3-second familiarization trials. The discrepancy between the results of the present study and those of Cantor and Cantor (1966) may be due to methodological differences between the

two studies, especially their use of a composite response measure. It is also possible that the results of the earlier study (Cantor and Cantor, 1966) may be attributable to chance.

The finding of an SFE for release but not for travel speed is in agreement with the results of other studies (Bogartz and Witte, 1966, Exp. II; Cantor and Cantor, 1964; Cantor and Fenson, 1968; Witte, 1965, 1967) which indicate that the SFE occurs in the initiation and not in the execution segment of the response when *Ss* are required to respond to stimulus onset.

A CE was found in the present study when release speeds were collapsed over four amount of familiarization groups (Groups 1, 10, 20, and 40). A release speed CE was also found for Group 0. A CE was also obtained when travel speeds were collapsed over Groups 1, 10, 20, and 40 but not for Group 0. Table 3 does show, however, that the difference between the travel speed means on change vs. nonchange trials for Group 0 *Ss* was in the same direction as for the other amount of familiarization groups—that is, speeds were faster on change than on nonchange trials. Thus, the children in this study started to respond and also completed the response faster on change than on nonchange trials.

Cantor and Cantor (1966) suggest that the CE may be due to the "novelty" inherent in changing from one stimulus to another from trial to trial within the motor task. The overall finding of a CE for both release and travel speeds (except in the case of travel speed for Group 0) is consistent with this hypothesis.

Bogartz and Witte (1966) suggest that a CE will be obtained only when *predominantly alternating* stimulus sequences are used in conjunction with a *differential* response. A CE for travel speeds was found in the present study although the stimulus sequences employed involved 50% alternation and 50% repetition of the NS and FS and thus were not predominantly alternating. These data indicate that it is not necessary to employ predominantly alternating stimulus sequences to obtain a differential response CE. A CE was also found for release speeds although this is not a differential response measure. This finding suggests that a CE can be obtained for a nondifferential initiation portion of a response if the execution portion involves a differential response.

A response interpretation involving assumptions different from those in the Bogartz and Witte (1966) hypothesis can account for the CE findings of the present study. One assumption, supported by some evidence (Craig and Myers, 1963; Jeffrey and Cohen, 1965; Manley and Miller, 1968; Weir, 1967), is that children in the age range employed in this study display strong response alternation tendencies in two-choice tasks. If these alternation tendencies either facilitate responding on change trials or in-

hibit responding on nonchange trials, then a response interpretation would account for the obtained travel speed CE. A second assumption is that, in addition to bringing strong alternation tendencies to the experiment, the children may have adopted a response strategy which involved waiting to release the start button until they had already decided which was the correct posterior button to push. In conjunction with this strategy, response alternation tendencies could also affect release speeds, either by increasing them on change trials, slowing them on nonchange trials, or a combination of both. Hence, a modified response interpretation can also account for the finding of a CE for the initiation portion of the response (i.e., release speeds).

REFERENCES

Bogartz, R. S., and Witte, K. L. On the locus of the stimulus familiarization effect in young children. *Journal of Experimental Child Psychology,* 1966, **4,** 317–331.

Cantor, G. N., and Cantor, J. H. Effects of conditioned-stimulus familiarization on instrumental learning in children. *Journal of Experimental Child Psychology,* 1964, **1,** 71–78.

Cantor, G. N., and Cantor, J. H. Discriminitive reaction time performance in preschool children as related to stimulus familiarization. *Journal of Experimental Child Psychology,* 1965, **2,** 1–9.

Cantor, G. N., and Cantor, J. H. Discriminitive reaction time in children as related to amount of stimulus familiarization. *Journal of Experimental Child Psychology,* 1966, **4,** 150–157.

Cantor, G. N., and Fenson, L. New data regarding the relationship of amount of familiarization to the stimulus familiarization effect (SFE). *Journal of Experimental Child Psychology,* 1968, **6,** 167–173.

Craig, G. J., and Myers, J. L. A developmental study of sequential two-choice decision making. *Child Development,* 1963, **34,** 483–493.

Jeffrey, W. E., and Cohen, L. B. Response tendencies of children in a two-choice situation. *Journal of Experimental Child Psychology,* 1965, **2,** 248–254.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Manley, S. K., and Miller, F. D. Factors affecting children's alternation and choice behaviors. *Psychonomic Science,* 1968, **13,** 65–66.

Miller, F. D., and Moffat, G. H. Stimulus intensity and familiarization effects on simple RT performance in children. *Psychonomic Science,* 1968, **13,** 289–290.

Weir, M. W. Children's behavior in probabilistic tasks. In W. W. Hartup and N. L. Smothergill (Eds.), *The young child.* Washington, D. C.: National Association for the Education of Young Children, 1967. Pp. 136–154.

Witte, K. L. Children's instrumental response speeds as related to conditioned-stimulus duration during a familiarization experience. Unpublished master's thesis, University of Iowa, 1965.

Witte, K. L. Children's response speeds to familiarized stimuli. *Psychonomic Science,* 1967, **7,** 153–154.

Witte, K. L., and Cantor, G. N. Children's response speeds to the offset of novel and familiar stimuli. *Journal of Experimental Child Psychology,* 1967, **5,** 372–380.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 118–126 (1969)

Imitation: Effects of Movement and Static Events

GREGORY T. FOUTS AND DAVID A. PARTON[1]
Institute of Child Behavior and Development, University of Iowa

In this study of imitation the modeling stimulus consists of the attributes of a model and the physical events produced by a model. The physical events include two components, the position of objects (static events) and the movement of objects (movement events). To investigate the effect of these components young children ($N = 64$) first observed one of four video tapes and then were given an opportunity to play with the materials seen on TV. The condition which presented a model and the condition which presented only the static and movement events produced similar amounts of imitation. Other results were that simple responses (reproduction of a single static event) required only the observation of static events; whereas complex responses (reproduction of two static events in the sequence presented by the modeling film) required the observation of static and movement events. The results suggest that a primary function of a model is the transmission of information about the environment.

The modeling stimulus used in most studies of imitation (Bandura, 1965) is a complex stimulus consisting of a human model manipulating objects. This study evaluates the influence of three components of a modeling stimulus on same responding in children. The three components are (*a*) the presence of a performing model, (*b*) static events, and (*c*) movement events. A static event is defined as a stimulus object at a specified location. A movement event is a stimulus object in transition from one location to another location. A same response is the occurrence of a static event by a subject. This definition is in terms of an apparatus change, not the topology of the subject's movement of the stimulus object.[2]

The effects that the three components of the modeling stimulus may

[1] This paper is based on a thesis submitted by the senior author in partial fulfillment of the requirements for the M.A. degree at the University of Iowa. The authors acknowledge the assistance of Richard A. Dubanoski and Kenneth Wissbrun and the cooperation of the staff of the Henry Sabin and William Penn schools of the Iowa City Community School District.

[2] The movement of a stimulus object by a subject can be viewed as a sequence of an infinite number of static events, and an operational definition of movement topology must be in terms of the timing and sequence of the locations assumed by the stimulus object. The stimulus object could be a portion of the subject's body, although inanimate stimulus objects are often used to achieve procedural simplicity in experiments.

have upon same responding are labeled the modeling effect, copying effect, and movement effect. To illustrate these effects, consider a situation in which marbles may fall into a transparent glass. If children drop more (or fewer) marbles into the glass after watching a model drop marbles than children who watched the marbles fall into the glass with no model present, then the difference in responding would be a modeling effect. If children drop more (or fewer) marbles into the glass after observing them fall into the glass in the absence of a model (i.e., after observing movement and static events) than children who saw only the marbles resting on the bottom of the glass (i.e., after observing only static events), then the difference would be a movement effect. Finally, if children drop more (or fewer) marbles into the glass after observing them rest on the bottom of the glass (i.e., after observing static events) than those children who saw no marbles in the glass (i.e., after observing the location), the difference would be a copying effect.

Two classes of same responses are examined in this paper, single same responses and same response chains. A single same response is the performance of a particular static event, and a same response chain is the performance of two static events in the order presented in the modeling stimulus. For example, a single same response could be the placement of a marble in a box. A same response chain could be the placement of a marble in a box followed by the placement of the same marble in a jar. The effect of the model, static events, and movement events on same responding should depend upon the complexity of the same response. The rationale for this expectation is presented below as each of the three effects is discussed.

Modeling effect. A recent investigation (Dubanoski and Parton, 1968) has shown that the observation of a human model is unnecessary for the subsequent emission of single same responses. They also found that a Model Hand condition, in which only the hand of a model was visible, produced the same amount of single same responses as a Model Absent condition in which stimulus objects were manipulated by invisible threads in the absence of a model. These findings suggest that a modeling effect is not a function of showing only the hand of a model, although it may be possible to produce a modeling effect if the hand had certain distinctive characteristics such as size or color. In the present study the model manipulation was a Hand condition in order to replicate this aspect of the previous study.

Movement effect. The observation of the movement event connecting two static events may serve an information function, e.g., the movement event demonstrates that a single stimulus object was transferred from one location to another. This information is likely to be irrelevant for the

production of single same responses but may be important in the production of response chains. If movement events facilitate the performance of response chains, then the Hand and Model Absent conditions should elicit more response chains than the Static condition. Such an outcome can be called a movement effect.

Copying effect. The contribution of static events can be determined by comparing performance under a Static condition with performance under a Location condition which presents only the locations at which the stimulus objects were placed in the Static condition. A Static condition should produce more single same responses than a Location condition because the locations alone provide no information regarding the potential locations specified for each stimulus object. A copying effect was not expected for response chains because both the Static and Location conditions lack the movement events which would seem necessary for the production of response chains.

METHOD

Subjects

Sixty-four first-grade children were randomly assigned to four experimental groups with the restriction that each group contain eight boys and eight girls.

Procedure

A male *E* escorted each *S* from the classroom to a research trailer. The *S* was seated 8 feet from a television set and instructed to watch the TV. While *S* watched a video tape, *E* sat behind him and inconspicuously timed the duration the *S*'s head was turned away from the TV screen.

Hand condition. In this video tape, the stimulus objects were manipulated by a young female adult with only the right hand and forearm of the model being visible on the television set. Three kinds of stimulus objects, 25 of each kind, were arranged in separate piles on a 22- × 48-inch table. The stimulus objects were small paper cups, large triangular paper clips, and 1.5-inch square wooden disks with holes in the centers. The stimulus objects were manipulated, one at a time, in the following sequences: (*a*) a cup was placed on a solid black 2.5-inch circle, then placed on one of 25 dowels which formed the branches of a 10.5-inch-high wooden "tree," and finally dropped into a transparent, 600-milliliter laboratory beaker; (*b*) a paper clip was placed on a black triangle, 1.0 inches on a side, placed edgewise in a white alley maze, and dropped into the beaker (the same beaker was used for all stimulus objects); (*c*) a disk was placed on one of nine 2-inch high square pegs mounted on a

board, placed onto one of three horizontal dowels protruding from a vertical board of triangular shape, and then dropped into the beaker.

Each sequence for a stimulus object lasted approximately 40 seconds and was presented three times, each time with another object of the same kind. Following the three cup sequences, the model performed three paper clip squences and three disk sequences. The model's hand was removed from the stimulus object at each location, and an object was visible at each location for two seconds. The stimulus object was removed from the beaker after the camera had focused on another part of the table. The seven locations and three piles of stimulus objects were arranged so that (*a*) no more than one of the 10 areas was visible on the television screen at any one time, and (*b*) the transitions from one location on the table to another did not expose any other location. The video tape was accompanied by symphonic music.

Model Absent condition. This video tape was identical to the Hand condition except that all the stimulus objects were manipulated by nylon threads which were not visible in the video presentation. No portion of the model was visible on the video tape.

Static condition. Each stimulus object manipulation in the Hand and Model Absent conditions involved two components, movement events and static events. The video tape for the Static condition was the same as for the Model Absent condition except that all movement events were deleted. The cup sequence, for example, showed a cup on the pile of cups for approximately two seconds followed by the camera scanning, with no cup present, the route leading to the circle for the same number of seconds as in the Model Absent condition. The cup, at the circle, was exposed for two seconds. This was followed by scanning to the tree, a 2-second presentation of the cup on the tree, scanning to the beaker, and a two-second presentation of the cup in the beaker.

Location condition. This video tape was identical to the Static condition tape except that there were no stimulus objects at any of the locations. This tape deleted both the movement events and static events. In the cup sequence, the pile of cups was presented for 2 seconds and this was followed by scanning to the circle, a 2-second exposure of the circle without a cup, scanning to the tree and a 2-second exposure of the tree without the cup, etc.

In summary, all video tapes were equated for the duration of exposure of each location, the duration of time used to scan from one location to another, the path the camera scanned, and the area of the table visible on the television screen.

Test for same responses. When the video tape ended, *E* pretended to adjust the television set, turned it off, and announced that he was going

into the next room of the trailer to fix the television wires. The *E* removed a screen which concealed the table and stimulus materials used in the video tapes. Two additional kinds of stimulus objects (black pegs and curtain rings) were included in order to decrease the likelihood of same responding occurring through the tendency of children to manipulate any objects presented to them. The *S* was seated facing the table and a one-way glass, and told that he could play with the toys in any way that he liked while *E* was trying to repair the television wires. The *E* entered the observation room where he served as the reliability observer during a 5-minute test for same responses. The *S*'s behavior was recorded continuously according to predetermined categories on a 20-channel event recorder. Following the test for same responding, the *E* returned to the experimental room and escorted *S* back to the classroom.

RESULTS

Each 5-minute record of behavior was scored to provide four dependent variables: (*a*) the number of different kinds of single same responses, with a range from zero to nine, (*b*) the total frequency of single same responses, collapsing over kinds of responses, (*c*) the number of different kinds of response chains, with a range from zero to six, and (*d*) the total frequency of response chains. The static events which were counted as components of a response chain were included in the count of single same responses.

Interobserver reliability for frequency of single same responses was obtained by dividing each record into consecutive 10-second intervals and assigning each *S* a score from each observer, the score being the count of the number of intervals in which at least one same response occurred. A product-moment correlation of .98 ($N = 45$) was obtained after eliminating all *S*s with scores of zero from both observers. The 10-second interval records were examined for interjudge agreement. Observer 2 recorded a same response on 96% of the 378 observation intervals on which Observer 1 recorded a same response. Observer 2 recorded the absence of a same response on 99% of the 1,392 intervals in which Observer 1 recorded the absence of a same response. There were 25 disagreements out of a possible 1,770 (59 *S*s $\times$ 30 Intervals; five *S*s were deleted from only this reliability analysis because of incomplete data for Observer 2 caused by an equipment failure). Out of the 25 disagreements, 11 apparently were timing errors; i.e., Observer 2 recorded a same response in either the interval preceding or following the interval in which Observer 1 recorded a same response, with the time between each pair of observations being less than 1 second.

Every analysis of single same responses which included the Location

group and every analysis of response chains was performed with median tests. The median test was used in these cases since only six of the 16 *Ss* in the Location group performed any single same responses. Similarly, only four of the 16 *Ss* in the Static condition and none in the Location condition exhibited response chains. Every statistic was considered significant if the null hypothesis was rejected at the .01 level for a two-tail test.

The data were first examined to determine whether same responding occurred without a human model performing the physical events. Table 1

TABLE 1

MEAN, MEDIAN AND *SD* FOR NUMBER OF DIFFERENT KINDS AND FREQUENCY OF SINGLE SAME RESPONSES AND SAME RESPONSE CHAINS

			Single Same Responses					
			Different Kinds			Frequency		
Condition	*n*	Prop. of subjects[a]	Mean	*SD*	Median	Mean	*SD*	Median
Hand	16	1.00	5.19	2.30	5.00	18.28	13.64	12.50
Model absent	16	0.88	4.31	2.05	4.83	11.06	7.60	9.00
Static	16	0.94	3.75	1.71	3.90	17.31	12.84	13.50
Location	16	0.38	0.69	1.04	0.30	5.06	10.69	0.30

			Same Response Chains					
			Different Kinds			Frequency		
Condition	*n*	Prop. of subjects[a]	Mean	*SD*	Median	Mean	*SD*	Median
Hand	16	0.56	1.31	1.76	0.75	2.50	5.11	0.75
Model absent	16	0.81	1.69	1.31	1.33	2.25	2.25	1.50
Static	16	0.25	0.25	0.44	0.17	0.25	0.44	0.17
Location	16	0.00	0.00	0.00	0.00	0.00	0.00	0.00

[a] Proportion of subjects with scores greater than zero (the same proportion necessarily applies to both Different Kinds and Frequency).

presents the means, *SD*'s and medians for each dependent variable according to experimental groups. The Location condition resulted in significantly fewer kinds and lower frequency of single same responses than each of the other conditions. The Hand, Model Absent, and Static conditions did not significantly differ in kinds or frequency of single same responses (F (2/42) = 2.07 and F (2/42) = 1.62, respectively). In addition, the Hand group did not differ from the Model Absent group

in performance of kinds or frequency of response chains. These results indicate that same responding can be obtained in the absence of a human model, a finding consistent with that of Dubanoski and Parton (1968). The absence of a modeling effect (i.e., a nonsignificant difference between the Hand and Model Absent conditions) suggests that merely providing the information that a human model is performing the movement and static events is not sufficient to produce a modeling effect.

The results provide support for a copying effect, i.e., the observation of static events was sufficient for the production of single same responses. The Static condition elicited significantly more single same responses than the Location condition, with 15 of the 16 *Ss* in the Static condition and six of the 16 *Ss* in the Location condition performing one or more same responses. The production of same responses by six of the *Ss* in the Location condition may have been due to that film presenting the correct locations immediately following the appropriate pile of stimulus objects. A second possibility is that the responding was due to one or more of the static events being a high probability response for young children in this situation, regardless of whether they have had prior exposure to the experimental situation.

In order to determine whether the observation of movement events influenced same responding, analyses were made with each dependent variable. The Hand and Model Absent groups exhibited significantly more kinds and greater frequency of response chains than the Location condition, and the Model Absent group performed significantly more kinds and greater frequency of response chains than the Static group. No significant differences were found for kinds or frequency of response chains between the Hand and Model Absent conditions, the Hand and Static conditions, or the Static and Location conditions. The differences between the Model Absent and Static conditions indicates that the performance of a response chain is a function, in part, of observing the movement event which intervenes between the two static events of the response chain. While a movement effect was obtained, it is not known what aspect of the movement manipulation was critical in producing the response chains. It is possible that the effective component of the movement event was simply the removal of the stimulus material from each location since that event could suggest that only one object was manipulated.

A movement effect was not found for single same responses; performances under the Hand, Model Absent, and Static conditions were not significantly different. This result indicates that single same responses can be elicited without the presentation of movement events. On the other hand, the presentation of movement events produced an increase in complex response chains. This suggests that when the modeling stimulus

involves simple manipulations of the environment, the occurrence of a movement effect depends on the complexity of the same response.

The *Ss*' inattention time to the television during the observation period varied from 0 to 155 seconds. Nearly half the *Ss* had zero inattention times and only two had inattention times of over 30 seconds. An extention of the Median χ^2 Test revealed no significant difference among the groups in inattention. The effect of sex of *S* on same responding was analyzed. Only one of the several overall and pair-wise analyses proved to be significant, the girls in an overall analysis performed more kinds of single same responses than boys, $F\ (1/42) = 6.66$.

The data were examined to determine whether the means and medians presented in Table 1 adequately represent performance under the various experimental conditions. Table 1 contains the proportion of subjects in each experimental group who exhibited one or more single same responses and one or more same response chains. Inspection of Table 1 reveals consistency across all three of the descriptive statistics, and an evaluation of the frequency distributions for each condition on each dependent variable indicates that these descriptive statistics adequately represent the data.

DISCUSSION

Investigations of imitation in humans occasionally assume that the reproduction of a model's behavior is a uniquely social phenomenon. That is, reproduction of the behaviors occurs as a function of observing another person producing changes in the environment. This study, and a prior study by Dubanoski and Parton (1968), found that providing the information that a human model performed the movement and static events does not produce a modeling effect. In view of the considerable amount of matching under the model condition, this result suggests that one function of a model is that of merely transmitting information concerning how the environment can be manipulated. It is not known, of course, to what extent the information transmission function of a model may account for the imitation effects which have been found in other experimental settings.

Qualifications of the above interpretation should be considered. First, it can be argued that *Ss* may covertly assume that the events in the Model Absent and Static conditions were produced by a human model and then respond accordingly. Such an interpretation may be impossible to investigate since it presumes events which at this time cannot be directly observed. A second argument is that this experimental analysis has limited scope because it does not allow for the study of behavior which matches the manipulations of only the model's body, e.g., a bare-

foot model marching briskly about a room. In this case, the static events and movement events, by definition, are inseparable from the model. The only logical implication of this criticism is that behavioral scientists may not be able to investigate the effects of the model independent of the information function when the behaviors are body movements.

The suggestion that a model serves the function of transmitting information concerning how the environment may be manipulated does *not* imply that there are no effects of model characteristics upon same responding. When the physical and/or behavioral characteristics of a model are distinctive, there are two possible effects. First, the characteristics may affect a subject's attention to the environmental changes produced by a model's behavior, thereby influencing the effectiveness of a model as a transmitter of events. Second, the characteristics of a model may elicit responses or processes which either suppress or facilitate the reproduction of the environmental events. The Model Absent condition may be used as one baseline for evaluating either of these effects of model characteristics.

Modeling stimuli, whether simple laboratory manipulations or complex behaviors, involve the three components investigated in this study. The effects of these components upon same responding label relations that may occur in any study of imitation. The relations demonstrated here may, therefore, account for imitation effects found in other experimental settings. Further analysis of the modeling stimulus is required in order to determine the generality of these three effects.

REFERENCES

BANDURA, A. Vicarious processes: A case of no-trial learning. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York: Academic Press, 1965. Pp. 1–55.

DUBANOSKI, R., AND PARTON, D. Imitation without a model. Paper presented at the Annual Meeting of the Eastern Psychological Association in Washington, D. C., April, 1968.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 127–139 (1969)

Attention, Anxiety, and Rules in Resistance-To-Deviation in Children

J. ALLAN CHEYNE, JOHN R. M. GOYECHE, AND RICHARD H. WALTERS[1]
University of Waterloo

An experiment was conducted which investigated the effects of punishment and rule instructions on resistance-to-deviation. Grade II boys were asked to select, over two trials, one of each of two pairs of toys. Following one selection subjects received a high-intensity (104 db) tone, a moderate-intensity (88 db) tone, or no tone. For half the subjects, instructions prohibiting touching the toy immediately followed this procedure. Measures of touching the selected toy in a subsequent test period in which the child was left alone with the toy suggested that, in the absence of instructions, subjects who received the high-intensity tone played least with the selected toy, whereas, among subjects receiving instructions, those under moderate-intensity punishment showed the least "deviant" play behavior. Supplementary data, including heart-rate monitored during training, suggested that under certain conditions, the role of attention in punishment training may be as important as that of "anxiety."

During recent years, the effectiveness of loud sounds as punishing stimuli has been demonstrated in a number of experimental studies employing child *Ss* (e.g. Aronfreed and Leff, 1963; Penney and Lupton, 1962; Walters and Parke, 1968). However, none of these studies has investigated the effects of varying the intensity of sound stimulation in the absence of prior prohibition instructions (cognitive structuring) that may facilitate, interfere with, or otherwise interact with the effects of manipulating the sound-intensity variable. For example, Cheyne and Walters (1969) have demonstrated that cognitive structuring may increase the effectiveness of low-intensity punishment for inducing response inhibition, particularly when punishment is delayed with respect to the prohibited response. On the other hand, there is evidence that such punishments may modify *Ss*' understanding or utilization of instructions.

[1]The authors express their gratitude for the co-operation of Superintendent, Principals, and Staffs of the Waterloo Separate School System. This study was supported by grants to the late Dr. R. H. Walters from the Ontario Mental Health Foundation, the National Institute of Health, United States Public Health Service and the Defense Research Board of Canada. This paper is dedicated to Dr. Walters who died during the preparation of this paper. Dr. Walters' enthusiasm and imagination were a source of inspiration for all who worked with him.

Cheyne and Walters' results indicate that punishment itself may, in some cases, attenuate the effects of cognitive structuring.

The experiment reported in this paper was designed to investigate some possible interactions between intensity of sound stimulation and the presence or absence of verbal instructions.

Learning theorists (e.g. Mowrer, 1960a,b) have suggested that, in punishment training, response-produced cues accompanying the punished act may become associated with anxiety-induced avoidance responses following punishment. Depending upon a number of parameters, such as the timing and intensity of punishment, the resultant association may lead to the subsequent inhibition of the punished act. The intensity of punishment is obviously an important factor since it will determine the strength of the avoidance response and hence the degree of inhibition of the punished response. Implicit in such theorizing is the assumption that the development of the association is relatively automatic. Little reference is made to the necessity of *Ss* attending to or being "aware" of the contingencies that are, in fact, being manipulated. On the basis of this theorizing it would be predicted that in the absence of verbal instructions there would be increasing response inhibition as intensity of punishment increases. Of course, there should be no response inhibition at those intensities which do not evoke anxiety responses. Strictly speaking, such a theory should properly be tested with humans in the absence of verbal cues, such as instructions, which may guide and direct *S*. As noted earlier, punishment studies involving children have consistently utilized verbal instructions prior to punishment training. In this study therefore, for all *Ss*, *prior* punishment instructions were eliminated, and for half of the *Ss*, *all* punishment-relevant instructions were eliminated.

Recent theorizing and research (e.g. Bandura and Walters, 1963; Easterbrook, 1959; Kausler and Trapp, 1960; Walters and Parke, 1964, 1965) have suggested that a moderate degree of emotional arousal induced by aversive physical stimuli may facilitate perceptual organization and the utilization of relevant cues. On the other hand, very intense aversive stimulation may have a disorganizing effect, resulting in distraction or lack of attention to relevant cues. Thus, there are grounds for investigating the relative roles of emotional arousal and attention in a punishment-training situation in which both the intensity of a punishment stimulus (noise) and the provision of verbal instructions immediately following punishment were varied.

Recent research concerning attentional factors has suggested that heart-rate deceleration is a correlate of attention to external stimuli, while heart-rate acceleration is associated with a reduction in sensitivity to external stimuli (Kagan and Lewis, 1965; Kagan and Roseman, 1964; Lacey, 1959, 1967; Lacey, Kagan, Lacey, and Moss, 1963; Lacey and

Lacey, 1964; Obrist, 1963). Graham and Clifton (1966) have recently extended the theoretical significance of cardiac activity by proposing that heart-rate deceleration is a component of the "orientation reaction" (Sokolov, 1963), while heart-rate acceleration is a component of the "defensive reaction" (Sokolov, 1963). Recent research, however, would indicate that the heart-rate reaction, during the orienting reaction, may be somewhat more complex than a simple linear deceleration (Greer, 1964; Germana and Klein, 1968; Goyeche and Thysell, 1969). In any case, it has been reported that the orienting response is accompanied by lowered sensory thresholds which facilitate the acceptance of stimulus input while the defensive or startle reaction is accompanied by heightened sensory thresholds which lead to the rejection of stimulus input (Lacey, 1967; Sokolov, 1963). Hence, it would be expected that attention to verbal cues would be facilitated under punishment conditions which elicit an orienting response and interfered with under punishment conditions which elicit a defensive or startle reaction. Cardiac activity was therefore monitored during punishment training in the present study as a possible index of attention.

On the basis of the above theorizing and research a number of predictions were made. Assuming that anxiety-mediation theory (e.g. Mowrer, 1960a) holds in the absence of verbal instructions, one would predict increasing resistance-to-deviation with increasing stimulus intensity in punishment training. However, considering the effectiveness of verbal instructions (Cheyne and Walters, 1969) in producing resistance-to-deviation, one would predict less deviation under conditions in which instructions are provided as opposed to those conditions in which instructions are not provided. Furthermore, it is predicted that under conditions in which instructions are provided, the learning of the instructions will be optimal under conditions of maximal attentional involvement. That is, a moderate-intensity buzzer that elicits the orienting response will facilitate learning of such instructions, and will result in superior resistance-to-deviation. On the other hand, a very high-intensity buzzer would be expected to elicit a defensive reaction, interfering with the learning of the instructions, thereby reducing the effectiveness of the instructions in producing resistance-to-deviation. Therefore, under instruction conditions the moderate-intensity condition should be more effective than the zero- or high-intensity conditions.

METHOD

Subjects

A total of 84 Grade II boys, for whom parental permission to participate was secured, were obtained from Kitchener-Waterloo separate

schools, and were randomly assigned to six different experimental conditions, 14 per group, involving three levels of intensity of physical punishment (0, 88, and 104 db) and the presence or absence of verbal instructions.

Procedure

Ss were tested in a mobile laboratory divided into an experimental and an observation room by a partition containing a one-way screen. Heart rate was recorded on a Beckman-Offner dynagraph by means of a telemetry system. Electrodes were connected to a miniature (1-1/2 × 1-1/2 × 1/2 inches) FM transmitter which transmitted the heart-rate signals via an FM receiver to a Beckman-Offner dynagraph located in the observation room.

As soon as *S* arrived at the mobile laboratory electrodes were attached to his sternum in order to secure continuous recording of heart rate. *E* then excused himself and went into the observation room to obtain a 1-minute basal heart-rate level. Upon his return, *E* then picked up two "unattractive" toys (two small, differently colored sports cars) from behind a cardboard partition on another table in the room. *E* placed the toys on the child's table (about 5 inches from the table's edge and about 10 inches apart); he asked *S* to "reach out and pick up the toy that you'd like to play with." *S*'s choice of either "unattractive" toy was never punished and both toys were removed 4 seconds after the toy was touched. *E* then presented two "attractive" toys (a green missle launcher and a white car with trailer) in the same positions repeating the same instructions. As the child touched either "attractive toy" he received a buzzer (the speaker being attached to the underside of the table) of either 0, 88, or 104 db (of 2-seconds duration) depending upon the experimental condition. Directly following this procedure *S* was immediately told either nothing, or, "Don't touch the white (green) toy, please," depending upon the experimental condition. The entire procedure from the moment the *S* touched one of the "attractive" toys, until the experimenter removed it, occupied a *total* of 4 seconds. Following the experimental manipulations, *E* placed the two previously chosen toys (the chosen "unattractive" toy which was never punished and the chosen "attractive" toy, which was punished in four of six conditions) on the table in front of *S* as if expecting him to make another choice. Before *S* could make his choice, *E* told *S*, "Oh no, I have forgotten something. I'm going to have to go into the school for a while. Will you be all right here by yourself? Good, I'll go out the door in the room next door. When I come back, I'll knock on the door so that you'll know that it's me. You can play here while I'm gone. O.K.?" *E* then left the experimental room,

and, after a moment, made a loud noise resembling that of the slammed trailer door. Actually, *E* remained in the observation room, stationed behind the one-way mirror in order to record *S*'s toy-playing activity during a 15-minute resistance-to-deviation period.

Measures

Resistance-to-deviation measures. During the 15-minute test period during which the child was left alone, the observer recorded, by means of a Heur-Century stop-watch, the times at which *S* touched and ceased to touch the chosen attractive toy. From these records the following measures were derived: The Latency of *S*'s first deviant response, the Number of times *S* touched the punished toy, and the total Duration of time during which *S* touched the punished toy. These measures are identical to those used in earlier research (e.g. Cheyne and Walters, 1969) and have been found to have interobserver reliability coefficients beyond .99.

Heart-rate measures. From the heart-rate recordings difference scores were determined over seven consecutive 2-second periods starting 2 seconds before the child's contact with the attractive toy. Difference scores were determined by substracting heart-rate scores during period one from each of the other six periods.

Supplementary measures. During the punishment training session, the direction of *S*'s looking responses during punishment training were tabulated by a naïve observer in an adjoining observation room according to which of the following targets he looked at: (1) Chosen attractive toy; (2) unchosen attractive toy; (3) experimenter; (4) elsewhere. The looking responses were recorded during the verbal instructions or an equivalent 2-second period for *S*s not receiving instructions.

Following the 15-minute resistance-to-deviation period, *E* re-entered the experimental room. *S* was then asked the following questions: (1) "Which toy did you play with while I was gone?" (2) "Did the buzzer frighten you at all?" (for the buzzer conditions). *E* merely recorded whether the child responded "Yes" or "No" to the questions and the child was thanked for helping *E* and was returned to the classroom.

RESULTS

Resistance-to-Deviation Data

Table 1 presents the group medians for Latency of the first deviation and the Number and Duration of deviations. Because of marked heterogeneity of variance and the skewness of the data distributions, a non-

TABLE 1
GROUP MEDIANS FOR THREE MEASURES OF RESISTANCE-TO-DEVIATION

	No instructions			Instructions		
	High	Moderate	Zero	High	Moderate	Zero
Latency	13.5	3.5	3.5	22.0	297.0	33.5
Number	10.0	10.5	12.5	6.5	2.5	7.0
Duration	123.5	379.0	396.0	123.5	17.0	104.5

parametric test, the Mann-Whitney U Test (Siegal, 1956), was used on all the behavioral data.[2]

To test the hypothesis that high-intensity punishment would be most effective under no-instruction conditions, scores for *Ss* under the high-intensity condition were compared with the scores of *Ss* under the combined conditions of moderate- and zero-intensity. The moderate- and zero-intensity conditions were also compared. The only comparisons that attained significance were the high-intensity condition versus the combined moderate- and zero-intensity conditions on the Latency measure ($U = 50$, $n_1 = 14$, $n_2 = 28$, $p < .0001$) and the Duration measure ($U = 60$, $n_1 = 14$, $n_2 = 28$, $p < .0001$). It may be seen from Table 1 that *Ss* under the high-intensity condition deviated later and for less time than *Ss* under the other two intensity conditions. The comparisons revealed no differences between *Ss* under moderate- and zero-intensity conditions on any of the measures.

In order to test the hypothesis that, under instruction-conditions, resistance-to-deviation would be greatest following moderate-intensity punishment, *Ss* under moderate-intensity-instruction conditions were compared to *Ss* under high- and low-intensity-instruction conditions. In addition, high- and low-intensity-instruction conditions were compared. Only the former comparison achieved significance. The moderate-intensity condition produced significantly longer Latencies of deviation ($U = 140$, $n_1 = 14$, $n_2 = 28$, $p < .06$), fewer deviations ($U = 121$, $n_1 = 14$, $n_2 = 28$, $p < .05$), and greater total deviations ($U = 150$, $n_1 = 14$, $n_2 = 28$, $p < .10$) than the combined high- and zero-intensity conditions.

In order to determine the effect of instructions at each intensity level, additional comparisons were made; namely, zero-intensity instructions versus zero-intensity–no-instructions, moderate-intensity instructions versus moderate-intensity–no instructions, and high-intensity instructions versus high-intensity–no-instructions. The comparisons yielded significant

[2] Parametric tests, however, yield results almost identical to those for the corresponding nonparametric tests reported here.

differences between instruction- and no-instruction groups under the zero-intensity condition on the Latency, Number, and Duration measures ($U = 10$, $n_1 = n_2 = 14$, $p < .001$), ($U = 49$, $n_1 = n_2 = 14$, $p < .05$), ($U = 45$, $n_1 = n_2 = 14$, $p < .01$), and under the moderate-intensity condition on the Latency, Number, and Duration measures ($U = 49$, $n_1 = n_2 = 14$, $p < .05$; $U = 59$, $n_1 = n_2 = 14$, $p < .05$; and $U = 46$, $n_1 = n_2 = 14$, $p < .01$). It may be seen from Table 1 that *Ss* under the moderate- and zero-intensity-instruction conditions deviated later, less often, and for less time than *Ss* under the moderate- and zero-intensity-no-instruction conditions.

Supplementary Data

Analysis of the observing responses of *Ss* during the 2-second instruction period revealed that *Ss* tended more often to look toward *E* and away from the toys under conditions involving the presentation of a buzzer (44 of 56 *Ss*) than under conditions without a buzzer (12 of 24 *Ss*) to a significant degree ($\chi^2 = 17.20$, $df = 2$, $p < .001$).

Analysis of the incidence of denial among *Ss* (question 1) indicated that deviating *Ss* under instruction conditions denied more often (8 of 36 deviators denying deviation) than deviating *Ss* under no-instruction conditions (1 of 40 deviators denying deviation) to a significant degree ($\chi^2 = 6.68$, $df = 1$, $p < .01$).

Significantly more *Ss* under the high-intensity condition verbalized negative reactions to the buzzer (responding "Yes" to question 2) than under the moderate-intensity condition ($\chi^2 = 5.73$, $df = 1$, $p < .02$). Under high-intensity conditions, 13 of 28 *Ss* verbalized negative reactions to the buzzer while under moderate-intensity conditions only 4 of 28 *Ss* did so.

Heart-Rate Data

Because of imperfect records, heart-rate data were available on only half the *Ss* (seven under each condition) during training. Figure 1 shows difference scores in beats per minute over seven 2-second periods. In Figure 1, period 1 represents 2 seconds immediately preceding the presentation of the punishing stimulus (or the equivalent for no-buzzer conditions); periods 2–7 represent subsequent 2-second periods. A summary of the analysis-of-variance of these data is presented in Table 2.

It may be seen from Table 2 that intensity significantly affected heart-rate difference scores. Figure 1 reveals that *Ss* under the high-intensity condition showed a greater increase in heart-rate scores than *Ss* under other conditions.

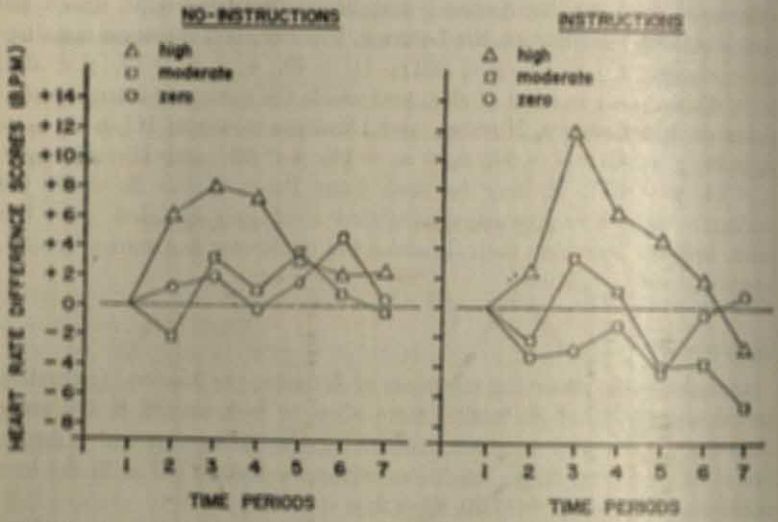


FIG. 1. Heart-rate change scores over seven 2-second periods for each of six experimental conditions.

Trend analyses revealed that only three groups demonstrated significant linear or quadratic trend components or both. Heart-rate change scores for *Ss* under high-intensity instructions showed both a linear trend ($F = 15.46$; $df = 1,78$; $p < .001$) and a quadratic trend ($F = 13.62$; $df = 1,78$; $p < .001$). Examination of Figure 1 reveals that the heart-rate for *Ss* under high-intensity conditions increased during periods 2 and 3 (during presentation of punishment and 2 seconds after) and decreased during the remaining periods, but only significantly so for the high-intensity-instruction group. A significant linear trend ($F = 17.38$; $df =$

TABLE 2
ANALYSIS-OF-VARIANCE OF HEART-RATE DATA

	SS	*df*	MS	*F*	*P*
Instructions	315.58	1	315.58	1.83	n.s.
Intensity	1136.59	2	568.29	3.30	.05
I × I	236.16	2	118.08	1	n.s.
Error (a)	6191.47	36	171.98		
Periods	728.62	5	145.72	6.08	.001
Instructions × periods	181.13	5	36.23	1.51	n.s.
Intensity × periods	1164.22	10	116.42	4.86	.001
I × I × P	115.30	10	11.53	1	n.s.
Error (b)	4311.30	180	23.95		

1,78; $p < .001$) for *Ss* under the moderate-intensity-instruction condition may be seen to be a consequence of the decrease in heart rate following period 2 (instruction period).

To further examine the relationship between heart-rate patterns during punishment training and resistance-to-deviation individual curves were plotted for high- and low-deviators (Figure 2), for moderate- and high-

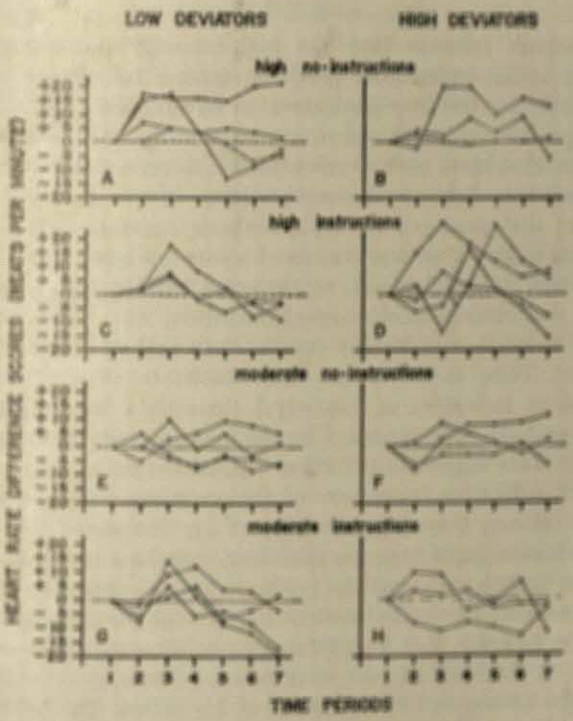


FIG. 2. Heart-rate change scores of seven 2-second periods for high- and low-deviators (graphs plotted for individual subjects).

intensity conditions. High and low deviators were *Ss* who were above or below the median, respectively, on at least two measures of resistance-to-deviation. It may be seen in Figure 2 that under no-instruction conditions, high- and low-deviators do not have discriminably different heart-rate patterns, under either high- (A and B) or low-intensity (E and F) conditions. Under instruction conditions, however, low deviators in both the high- and moderate-intensity groups appear to have characteristic patterns. The pattern for *Ss* under high-intensity conditions (C) is bi-

phasic, first increasing following the buzzer presentation and then decreasing after the instruction period. Under moderate-intensity conditions (E), *S*'s heart-rate patterns may be seen to be essentially triphasic, an initial deceleration followed by an acceleration, which in turn is followed by a deceleration. No such consistent pattern may be discerned for high-deviators under high- or low-intensity-instruction conditions (D and H).

DISCUSSION

The analyses indicate that the high-intensity punishment in the absence of verbal instructions produced greater resistance-to-deviation than the moderate-intensity punishment or no punishment. The moderate-intensity punishment, furthermore, produced no more inhibition than no punishment when there were no subsequent instructions given.

The heart-rate data are instructive here. It was the high-intensity punishment that produced a simple biphasic response, that is, marked acceleration followed by deceleration of somewhat lesser magnitude. The moderate-intensity punishment, on the other hand, produced a triphasic response, deceleration-acceleration-deceleration, the accelerative component of which is considerably smaller than that of the high-intensity punishment. Thus, to the extent that heart-rate acceleration may be interpreted as indicative of "anxiety," these data would suggest that anxiety may have been produced by the high-intensity punishment condition *only*. This suggestion is further supported by *S*'s reports (question 2) and *E*'s subjective impression of the aversiveness of both stimulus intensities. Hence, there may be support for the suggestion (Mowrer, 1960a) that subsequent response inhibition may be a result of "anxiety" or aversive "emotional" reactions produced by high-intensity punishment in the absence of any verbal structuring through instructions. It is also conceivable, however, that high-intensity punishments are more effective cues that, on the basis of past learning experiences, enable the child to evaluate the nature and consequences of his action. The heart-rate acceleration data would tend to support the anxiety mediated theory, although the possibility of cognitively controlled autonomic responses should not be overlooked.

On the other hand, the moderate-intensity buzzer produced a triphasic pattern, the initial component of which was decelerative. It has been suggested that this deceleration represents one component of the orienting response (Graham and Clifton, 1966). There is, however, evidence that the cardiac response accompanying the orienting reaction is somewhat more complex than this (Geer, 1964; Germana and Klein, 1968). In a recent study of the orienting response in a "signal" situation (Goyeche and Thysell, 1969) a triphasic response (deceleration-acceleration-de-

celeration) was found. Hence, the entire cardiac pattern of *Ss* in the moderate-intensity groups may be indicative of an orienting response. Thus, it is suggested that the superior resistance-to-deviation of *Ss* who had received the moderate-intensity buzzer followed by instructions may be attributed to their greater attention to, and comprehension of, the instructions.

The immediately preceding discussion has made the inference that those *Ss* who were low deviators under moderate-intensity instructions were those who displayed orienting responses as indicated by the heart-rate pattern. Indeed, the heart-rate patterns of low and high individual deviators under moderate-intensity instruction (Figure 2) suggests that this assumption is not without support. Low deviators in the moderate-intensity-instruction group demonstrated the previously mentioned triphasic response, while high deviators appeared to show a random response.

Under conditions in which instructions followed punishment, the moderate-intensity punishment appeared to be the most effective condition. Additional comparisons revealed that both moderate- and zero-intensity punishment were significantly more effective when accompanied by instructions, on all measures of response inhibition, while high-intensity punishment was no more effective when accompanied by instructions. This finding combined with the fact that the most effective overall condition was the moderate-intensity instruction condition suggests that the high-intensity punishment did not facilitate the comprehension and/or utilization of the additional information supplied by the instructions. Since the high-intensity condition produced a marked heart-rate acceleration which has, in the literature, been associated with a defensive reaction (Graham and Clifton, 1966) and heightened sensory thresholds (e.g. Lacey and Lacey, 1964), some interference with comprehension of the instructions would be expected. It is interesting that subjects under high- and moderate-intensity-instruction conditions showed marked deceleration following the instructions (Figure 1). It may further be noted that the low-deviating subjects under both high- and low-intensity instruction conditions were, in fact, the subjects who showed a heart-rate deceleration following the instruction period (Figure 2,C and G). High deviators under these conditions showed no consistent heart-rate pattern (Figure 2,D and H).

It may be suggested then that in the absence of preliminary instructions, the buzzers produced a great deal of uncertainty regarding the reasons for the onset of the noise. It may be recalled at this point that significantly more subjects who received both moderate- and high-intensity buzzers tended to look to the experimenter than under no-

buzzer conditions. This could be interpreted as the subjects' seeking information from the experimenter concerning the meaning of the buzzer. The instructions immediately following the buzzer may have served to reduce uncertainty in those groups of subjects in which the late decelerative component and greater subsequent resistance-to-deviation was present.

In summary, those subjects in the instruction conditions which showed the greatest resistance-to-deviation could be identified by two cardiac characteristics: (1) A triphasic H.R. response following an 88-db sound, which is interpreted as an index of the O.R.; (2) an augmentation of the late decelerative component following instructions which is interpreted as reflecting a reduction of uncertainty.

The fact that 22% of deviating subjects receiving the instructions denied deviation while only 2.5% of subjects receiving no instructions denied deviation agrees with findings from previous studies (Andres, 1967; Cheyne and Walters, 1969). This denial can probably be attributed to a clearer knowledge of the prohibition by subjects receiving added "rule structure" through instructions (Cheyne and Walters, 1969).

Kagan has argued that "the essence of learning is more dependent on attentional involvement by the learner than on specific quantities of particular external events" (Kagan, 1967, p. 134). More specifically, the findings of this experiment indicate that, in the case of punishment training with children in which verbal instructions follow, as is often the case under natural conditions, the efficacy of learning depends at least as much upon those properties of stimulus events which elicit attention as those properties which produce "anxiety."

REFERENCES

ANDRES, D. H. Modification of delay-of-punishment effects through cognitive restructuring. Unpublished doctoral dissertation, University of Waterloo, 1967.

ARONFREED, J., AND LEFF, R. The effects of intensity of punishment and complexity of discrimination upon the generalization of an internalized inhibition. Unpublished manuscript, University of Pennsylvania, 1963.

BANDURA, A., AND WALTERS, R. H. *Social learning and personality development.* New York: Holt, 1963.

CHEYNE, J. A., AND WALTERS, R. H. Intensity of punishment, timing of punishment, and cognitive structure as determinants of response inhibition. *Journal of Experimental Child Psychology,* 1969, **7**, 231–244.

EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychological Review,* 1959, **66**, 183–201.

GEER, J. H. Measurement of the conditioned cardiac response. *Journal of Comparative and Physiological Psychology,* 1964, **57**, 426–433.

GERMANA, J., AND KLEIN, S. B. The cardiac component of the orienting response. *Psychophysiology,* 1968, **4**, 324–328.

GOYECHE, J. R. M., AND THYSELL, R. V. The cardiac index of the O.R. as a function of anticipation interval. Submitted for publication, 1969.

GRAHAM, FRANCES, K., AND CLIFTON, RACHEL K. Heart-rate change as a component of the orienting response. *Psychological Bulletin,* 1966, **65,** 305–320.

KAGAN, J. On the need for relativism. *American Psychologist,* 1967, **2,** 131–142.

KAGAN, J., AND LEWIS, M. Studies of attention in the human infant. *Merrill-Palmer Quarterly,* 1965, **11,** 95–127.

KAGAN, J., AND ROSEMAN, B. L. Cardiac and respiratory correlates of attention and an analytic attitude. *Journal of Experimental Child Psychology,* 1964, **1,** 50–63.

KAUSLER, D. H., AND TRAPP, E. P. Motivation and cue utilization in intentional and incidental learning. *Psychological Review,* 1960, **67,** 373–379.

LACEY, B., AND LACEY, J. I. Cardiac deceleration and simple visual reaction time in a fixed foreperiod experiment. Paper read at the Society for Psychophysiological Research, Washington, D. C., October, 1964.

LACEY, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic processes and outcome. In E. A. Rubenstein and M. B. Parloff (Eds.), *Research in psychotherapy.* Washington, D. C.: American Psychological Association, 1959, pp. 160–208.

LACEY, J. I. Somatic response patterning and stress: some revisions of activation theory. In M. H. Appley and R. Turnbull (Eds.), *Psychological stress: issues in research.* New York: Appleton, 1967.

LACEY, J. I., KAGAN, J., LACEY, B. C., AND MOSS, H. A. The visceral level: situational determinants and behavioral correlates of autonomic response patterns. In P. Knapp (Ed.), *Expression of emotions in man.* New York: International Universities Press, 1963, pp. 161–196.

MOWRER, O. H. *Learning theory and behavior.* New York: Wiley, 1960a.

MOWRER, O. H. *Learning theory and the symbolic processes.* New York: Wiley, 1960b.

OBRIST, P. A. Cardiovascular differentiation of sensory stimuli. *Psychosomatic Medicine,* 1963, **25,** 450–459.

PARKE, R. D., AND WALTERS, R. H. Some factors influencing the efficacy of punishment training for inducing response inhibition. *Monographs of the Society for Research in Child Development,* 1967, **32,** No. 1 (Serial No. 109).

PENNEY, R. K., AND LUPTON, A. A. Children's discrimination learning as a function of reward and punishment. *Journal of Comparative and Physiological Psychology,* 1962, **65,** 351–354.

SIEGAL, S. *Nonparametric statistics for the behavioural sciences.* New York: McGraw-Hill, 1956.

SOKOLOV, E. N. *Perception and the conditioned reflex.* Oxford: Pergamon, 1963.

WALTERS, R. H., AND PARKE, R. D. Social motivation, dependency, and susceptibility to social influence. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 1. New York: Academic Press, 1964, pp. 231–276.

WALTERS, R. H., AND PARKE, R. D. The role of the distance receptors in the development of social responsiveness. In L. P. Lipsett and R. D. Spiker (Eds.), *Advances in child development and behavior.* Vol. 2. New York: Academic Press, 1965, pp. 59–96.

WALTERS, R. H., AND PARKE, R. D. The influence of punishment and related disciplinary techniques on the social behavior of children: theory and empirical findings. In B. A. Maher (Ed.), *Progress in experimental personality research.* Vol. 3. New York: Academic Press, 1968, pp. 179–228.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 140–146 (1969)

Positive Reinforcing Function of "Negative Attention"[1]

RONALD GALLIMORE AND ROLAND G. THARP
University of Hawaii

AND

BRYAN KEMP
University of Southern California

It has often been suggested that some children, in the absence of positive attention, prefer negative attention to indifference. To test this proposition experimentally, twenty-six third and fourth grade children were randomly assigned to (1) a period of social deprivation or (2) a brief, but warm interaction with the experimenter. All subjects subsequently played a game-like probability matching task in which face to face contact and a mild reproof were contingent upon pressing a button which had little likelihood of being "right." Given the choice between being "right" (seeking a light flash) and being "wrong" (seeing the experimenter's face and hearing him say, "you're wrong"), socially deprived, high need for approval subjects showed a significantly greater tendency than controls to choose the latter.

Some children appear to consistently behave in ways which elicit adult disapproval. Such children are sometimes characterized as "negative-attention seeking"; some observers have gone so far as to suggest that such individuals exhibit a "need" to be punished (Josselyn, 1948) while others have argued that children simply need attention, even if it is negative (Davis, 1949). Social learning theorists, on the other hand, assume that what an adult may regard as negative-attention seeking is, in many instances, the consequence of an intermittent reward and punishment schedule which is not apparent to the observer (Bandura and Walters, 1963). That is, the observer incorrectly infers that a child *intends* to provoke negative attention. To illustrate, a pre-schooler may persist in disrupting the teacher-led activities of his classmates despite the fact that

[1] This work was partially supported by a faculty research grant from the Foundation of the California State College, Long Beach, California. Acknowledgment is due John E. Nygaard for his contributions to the experimental design, and to Lee Sechrest and Edward Kubany for several important suggestions. Please address inquiries to R. Gallimore, Social Science Research Institute, University of Hawaii, Honolulu, Hawaii 96822.

he is often reprimanded for such behavior—that which maintains the behavior are the reinforcing events which occur, perhaps while the teacher is absent or distracted, when his antics provoke a positive peer-response. Thus it is unnecessary to assume that the child either needs or enjoys the negative reaction of his teacher. Indeed, the use of the term in such cases is descriptively ambiguous, and theoretically unnecessary.

Bandura and Walters (1963) also suggest that many responses labeled negative-attention seeking may be examples of anxiety-motivated responses which are often highly resistant to extinction. An individual may persistently expose himself to aversive, or negative attention, in his efforts to escape more unpleasant consequences. Such a conception has been, of course, widely accepted as an explanation for the persistence of self-defeating, neurotic behavior. Both explanations, attributed above to Bandura and Walters, describe circumstances in which attention only *seems* to be desirable: yet it would have no positive-reinforcing function of its own.

Despite wide interest in negative-attention as a descriptive category, as an explanation for irritating behavior, and as a problem for conceptual analysis, there apparently have been no published attempts to experimentally produce the phenomenon. Thomas, Becker, and Armstrong (1968) found that when systematically varying teacher behavior from "approving" behavior to "disapproving," the frequencies of student disruptiveness also rose; and there is evidence that some forms of critical comment do strengthen the antecedent behavior (if children stand up, and the teacher says "sit down," they stand up the more; Madsen *et al.*, 1968). The present study is an attempt to expand the inquiry concerning the function of "negative-attention" to the laboratory. In order to provide some experimental data on these phenomena, the work reported here was designed to ask a relatively simple empirical question—if children in a novel situation are offered a choice between (1) making a correct response (signaled in an impersonal fashion) which offers no adult attention and (2) making an incorrect response which elicits mild adult reproof, are there any circumstances under which they will select the latter?

Such circumstances might be expected to involve several factors, two of which were selected for experimental manipulation: *social deprivation* and *relative reinforcing strength of social approval.* This latter factor has been more often described as "need for social approval"; as such, both it and social deprivation have been demonstrated as related to the arousal of "attending and orienting responses" (Crowne and Marlowe, 1964; Walters and Parke, 1964). Studies in the tradition of Gewirtz and Baer (1958) suggest that children are trained to attend and orient to adults following social deprivation; and Crowne and Marlowe (op. cit.)

appear to conclude that the disposition to seek social approval reflects a history of reward for attending and orienting to adult-produced cues. We might thus expect that the interaction of these two variables may produce, in the absence of positive attention, a preference for negative attention (mild adult reproof) to no attention.

In other language, but using identical operations, mild adult reproof may in some circumstances for some children function as a positive reinforcer, even for undesirable behavior. We sought to produce, in a contrived setting, a response analogue of this phenomenon, so frequently reported in clinical, applied, and lay settings.

METHOD

Overview

Twenty-six third and fourth grade boys and girls were randomly assigned to (1) a period of social deprivation or (2) a brief, but warm interaction with the experimenter (*E*). All subjects (*Ss*) subsequently played a game-like probability matching task in which face to face contact and a negative verbal response by the *E* were contingent upon the perseveration of an incorrect response.

Measure of Reinforcing Value of Social Approval

The Children's Social Desirability scale (*CSD*) (Crandall and Katkovsky, 1965) was selected as the measure of need for social approval: This questionnaire is similar in construction to the adult scale developed by Crowne and Marlowe (op. cit.). Initial findings suggest strongly that the *CSD* measures the same dimension among children as that described by Crowne and Marlowe for adults. We assumed in using the *CSD*, that the characterization of high-scoring adults (those who subscribe to socially desirable, but improbably flattering self-descriptions) presented by Crowne and Marlowe could be generalized to children. We assumed that *CSD* scores reflect, in part, the disposition to *attend* to others in the pursuit of social approval, as well as the probable reinforcing value of that approval.

Apparatus

The apparatus, similar to that described by Lewis, Wall, and Aronfreed (1963), consisted of a panel with a start light, two buttons, and two signal bulbs. The *E* initiated each trial by activating the start light to which the *S* responded by depressing one of two buttons. Correct choices were indicated by the signal bulb directly above the selected button: errors activated the bulb above the adjacent bulb. Frequency of choice for each button was recorded automatically.

The game apparatus was mounted in a large cardboard booth: a sliding panel above the game permitted *E* to look into the booth and speak with *S* at any time. The *S* had visual contact with *E* only when the panel was open.

Procedure

Social deprivation. As each child was excused by his teacher he was told "go with this man"; no further explanation was offered. After escorting the *S* to the experimental room, *E* explained that he was from a toy company and wanted to test a new game. Otherwise, the *E* remained "aloof," initiated no conversation, and responded to questions with an unelaborated "yes" or "no."

At this point, half the *Ss* were given the social desirability scale and half were given the task instructions. The *E* then said he must return a call to his company; the *S* was instructed to sit inside the booth, with the door closed and not to touch anything or leave the booth until *E* returned. Approximately 10 minutes later, *E* returned (gave task instructions to the remaining half of the *Ss*) and the training commenced. At the end of the training trials, those *Ss* who had not yet completed the *CSD*, then did so.

Social satiation. As in the deprivation condition, *Ss* were dismissed with no explanation, but in the social satiation condition, *E* was friendly and talkative, offering, for example, a drink from a fountain during the walk to the experimental room. Next, *E* explained the task (to half these satiation subjects) prior to a 10-minute conversation about school and topics of interest to the *S*: the remaining half of the *Ss* received the instructions after the 10-minute interaction. *CSD* administration was managed exactly as for the socially deprived group, one-half before and one-half after the training trials.

The tactic of randomly administering the task instructions before and after the trials was chosen to avoid unwanted experimenter effects on the satiation-deprivation condition; for example, if all children had been given instructions following the deprivation condition, this social contact might have reduced the potency of the deprivation. In a parallel way, the satiation condition might have been weakened, had the "impersonal" instructions followed the "satiating" interaction.

The *CSD* was also administered in premanipulation and posttraining trials-conditions, one randomly chosen half of the subjects in each condition. This follows the procedure employed by Crowne and Marlowe in their work with adults. It is designed to avoid reciprocal contamination of the scale and the experimental task.

Instructions. Ss were told, in essence, to push one button each time the red start light came on and that the object of the game was to match the

button and the right-wrong signal lamp as often as possible. It was made clear that either button could be correct, but only one on any given trial. *S* was told to play until *E* said to stop.

Negative attention. Each *S* completed 100 training trials. Events were randomized within each block of ten trials on an 80% to 20% ratio of reinforcement with alternation of the contingencies across *S*s to reduce right-left bias. Throughout the training trials, *E* remained outside the booth. Each time *S* selected the low probability choice (*R*.20) and was wrong, *E* opened the sliding panel, looked at the *S* and said, "no, you're wrong." In other words, if the *R*.20 choice was selected but the *R*.80 choice was correct, *S* received the attention of *E*. At all other times, correctness and incorrectness were indicated by the signal lights on the game panel. Care was taken to avoid making the brief contact highly aversive, but at the same time it was made clear to the *S* that he had been wrong. Telling the *S* he is wrong does not, of course, exhaust the possible operational definitions of negative attention. In this instance it was anticipated that a relatively mild, as opposed to severe, negative comment would be more congruent with the potency level of the experimental manipulations.

Dependent measure. Frequency of response to the *R*.80 side constituted the dependent variable. It was predicted that high need approval *S*s in the deprived condition would match the contingent probabilities less rapidly and less well than the control groups.

RESULTS

Subjects were divided at the overall *CSD* median, yielding four groups: high *CSD*-deprived, high *CSD*-satiated, low *CSD*-deprived, and low *CSD*-satiated.

Figure 1 presents the mean proportion of responses to the *R*.80 side, with the data divided into five blocks of 20 trials each. As predicted, the deprived high social-reinforcer group exhibits fewer responses to the high probability side than the other three groups, with the largest difference occurring on the final 20 trials.

A Kruskal-Wallis one-way analysis of variance by ranks indicated that the variation in ranks among the four groups differed significantly across the five trial blocks ($p < .05$), in terms of frequency of response to the *R*.80 side. Testing for differences among groups on the final block of 20 trials with the Mann-Whitney *U* test reveals the high-deprived group to be significantly less likely than the other three groups to select the high probability side. This analysis of the final trial block, presented in Table 1, also indicates no significant difference among the other three groups. Similar analyses of trial blocks three and four yielded either marginally significant differences ($p < .10$) or nonsignificant differences. While the

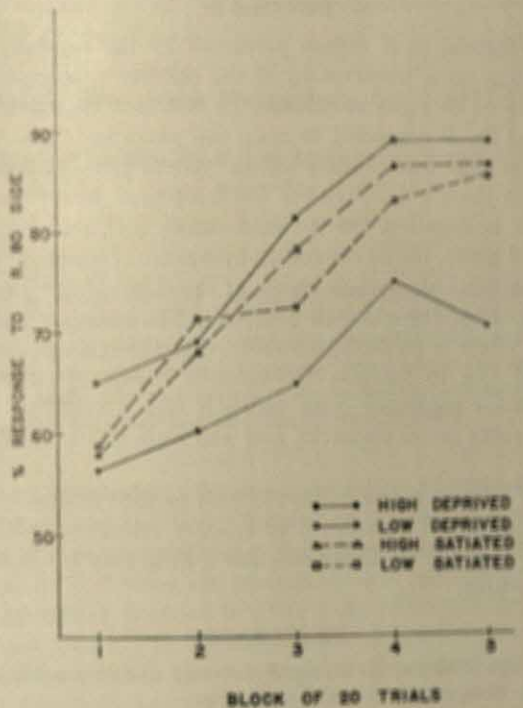


FIG. 1. Mean Proportion of response to the more frequently reinforced side.

results of the Kruskal-Wallis test and curves in Figure 1 suggest the high-deprived *S*s are consistently and significantly less likely to select the *R*.80 choice, the difference between their performance and the three remaining groups becomes pronounced only on the final 20 trials.

TABLE 1
COMPARISONS OF EXPERIMENTAL CONDITIONS ON FINAL TWENTY TRIALS

Trials	Hi *n* App Deprived ($N = 6$)	Hi *n* App Satiated	Lo *n* App Satiated
81-100			
Hi *n* App Satiated ($N = 6$)	.001*	—	—
Lo *n* App Satiated ($N = 7$)	.05	*NS*	—
Lo *n* App ($N = 7$)	.01	*NS*	*NS*

* *p* Values.

DISCUSSION

Was the response of *E* indeed *perceived* by the children as negative attention? In general, observation of the children's behavior in all conditions suggested no signs of distress. In fact on Trial 97 for one of the high-deprived *S*s, *E* neglected to open the panel and offer the negative comment—the child immediately asked, "Aren't you going to tell me I'm wrong?" Thus there is no basis for assuming that the children perceived *E*'s comment as negative. On the other hand, with the exception of the high-deprived group, the *S*s showed an avoidance of the negative comment —by the final trial block their choice of the *R*.80 side averaged approximately 90%. How the children perceived *E*'s comment is a matter of speculation. There is no doubt, however, that the functional value of the mild reproof was to increase the relative frequency of those clearly labeled incorrect responses upon which it was contingent, for children who highly value social approval, and who were mildly deprived of that approval.

Of further interest is the support offered to the *CSD* as an instrument for at least a partial assessment of relative reinforcing value of social approval, and thus of the likelihood that a child may be "mis-taught" by negative-attention.

REFERENCES

BANDURA, A., AND WALTERS, R. H. *Social Learning and Personality Development.* New York: Holt, Rinehart, and Winston, 1963.

CRANDALL, V. C., CRANDALL, V. J., AND KATKOVSKY, W. A children's social desirability questionnaire. *Journal of Consulting Psychology,* 1965, **29,** 27–36.

CROWNE, D. P., AND MARLOWE, D. *The Approval Motive.* New York: Wiley, 1964.

DAVIS, K. *Human Society.* New York: Macmillan, 1949.

GEWIRTZ, J. L., AND BAER, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *Journal of Abnormal and Social Psychology,* 1958, **56,** 165–172.

JOSSELYN, I. M. *Psychosocial Development of Children.* New York: Family Service Assoc. of America, 1948.

LEWIS, M., WALL, A. M., AND ARONFREED, J. Developmental change in the relative values of social and nonsocial reinforcement. *Journal of Experimental Psychology.* 1963, **66,** 133–137.

MADSEN, C. H., JR., BECKER, W. C., THOMAS, D. R., KOSER, LINDA, AND PLAGER, ELAINE. An analysis of the reinforcing function of "sit down" commands. In R. K. Parker (Ed.), *Readings in Educational Psychology.* Boston: Allyn and Bacon, 1968.

THOMAS, DON R., BECKER, WESLEY C., AND ARMSTRONG, MARIANNE. Production and elimination of disruptive classroom behavior by systematically varying teacher's behavior. *Journal of Applied Behavior Analysis,* 1968, **1,** 35–46.

WALTERS, R. H., AND PARKE, R. D. Social motivation dependency, and susceptibility to social influence. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology, I.* New York: Academic Press, 1964, pp. 231–276.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 147–159 (1969)

Field Trials of a New Procedure for Toilet Training[1]

R. KEITH VAN WAGENEN, LEE MEYERSON, NANCY J. KERR, AND KURT MAHONEY
Arizona State University

A new procedure for toilet training, utilizing the discrimination and reinforcement opportunities of a rapid, forward-moving series of training events, successfully trained nine profoundly retarded children. Each training sequence was initiated by an auditory signal which was activated by the flow of urine. The new behavior was successfully maintained in other environments when access to a toilet facility was continued.

The culture has stringent requirements for the disposal of body waste and, in consequence, the toilet training procedures of mothers may be coercive, restrictive, and punishing. Perhaps one means by which children could be spared the stresses, and what are said to be life-long consequences which are sometimes assumed to accompany toilet training, would be to make the task easier. Paradoxically, however, almost no systematic attention has been paid to the devising of schema for improving the efficiency of the task, i.e., facilitating the formation of the necessary discriminations in children.

An almost universal procedure for training involves the placing of the child on a pot and waiting for a toileting event to occur. The obvious basis for the method is that without potting the voiding event is usually private or concealed by clothing, thus providing no way for the mother to effect a training regimen. Successful training at an early age, therefore, is dependent first on the coercive restraint of locomotion at the commode while the mother observes, and later in the training when the child is not potted, on the mother's high vigilance and sensitivity to subtle cues from the child. The mother requires these cues to indicate when she should rush her child to the facility, so that he may learn to anticipate his own requirement and thenceforth take himself to the commode.

Success in training is adversely affected by the large segments of time separating practice of one element of the behavior from practice on other elements and by the order of the training events. The training trials of

[1]This investigation was aided, in part, by OE Project 5-0415 and by the Mesa Association for Retarded Children School, Mesa, Arizona. Assistance of the Directress, Mrs. Darlene Schaumburg, is gratefully recognized.

the potting procedure, whether bowel or bladder training, do not lead to practice of the toileting behavior in the same temporal sequence as will be required when the training is complete, but will train one or two elements of the chain at a time beginning at the terminal end and working backward as follows: (1) Potting to control the locus of the final event with limited attention to releasing and initiating sphincter control, the "turn on-turn off" behavior. (2) Bringing about restraint of voiding at all other locations. (3) Inducing movement to the commode when a voiding event has begun or appears imminent. (4) Teaching the child to manage clothing. (5) Teaching the *S* to take an appropriate final stance.

The time elapsed between the first training element and the last element is likely to be weeks or months unless training has been deferred until the child has sufficient verbal repertory that instructions may be given to him. In retrospect, the limitations of the maternal training procedure seem evident, but they also appear to be surprisingly difficult to overcome. Current popular literature (Spock, 1968) and even the scientific or relatively systematic literature consists primarily of procedures that begin with potting (Cassell and Colwell, 1965; Dayan, 1964; Ellis, 1963; Giles and Wolf, 1966; Hunziak *et al.*, 1965; McGraw, 1940; Pumroy and Pumroy, 1965).

METHOD

A different procedure. This paper reports field training trials of a new procedure for toilet training that changes the emphasis. Rather than placing the child on a pot when immediate voiding is unlikely, it appears reasonable that the learning to initiate urination, in the proper receptacle, could be learned best at those times when reflex voiding occurs. Training efforts will be enhanced if this reflex occurs in contiguous temporal relationship to environmental events designed to bring urination under control. This may be accomplished by equipping the trainee with a signal device which sounds an alert when urination begins. With the use of such a device the trainee is free to move about until a signal announces that urination has begun, at which time the trainer may intervene to establish appropriate learning. Each training trial then takes the form of a quick sequence of events which closely approximate the desired terminal behavior. The sequence is as follows: (1) Each time urination begins, an auditory alert sounds and the trainer makes a rapid approach to the *S*, calling "no." The frequent effect on the subject is termination of urine flow (a mild startle effect). (2) The trainer grasps the *S* by the hand and moves him quickly toward the toilet facility. (3) Places the *S* at the commode in an appropriate final stance for voiding. (4) Urination is reinitiated or continued, as the case may be, under conditions of quiet and

the *S* is approved. (5) After a variable number of trials, the *S* emits the appropriate sequence of toilet behaviors preliminary to voiding, and without the intervention of the trainer, the auditory signal device is replaced with cotton training briefs and the trainer teaches the *S* to remove and replace the pants at the appropriate times.

To distinguish this training procedure from the potting procedure, it will be referred to hereafter as "the forward-moving" sequence. Elapsed time in the forward-moving sequence is a matter of a few seconds, for in this approach the passive potting procedure is replaced by rapid actions of a trainer moving the *S* quickly in performance of the necessary behaviors. Each trial includes practice on most of the elements comprising the terminal behavior. Step 5 above is an exception as it is not included in the early trials.

Before the trainee can be equipped with the usual undergarments, an aspect of step five, he is required to learn to respond appropriately without the intervention of the trainer. When the *S* advances to this degree of independence, the trainer "fades out" his approach to the *S*; at first, approaching close while beckoning with the hands and calling to the *S* in order to induce walking toward the facility. These prompts are removed as quickly as the behavior permits. Independent behavior of the subject usually requires but a few days, at which time the final element of training, clothing manipulation, may be undertaken. Teaching a nonverbal human to remove and replace clothing is a difficult task worthy of independent investigation. It has been accomplished satisfactorily in this toilet training research by superimposing the trainer's hands over those of the subject, making the required movements, and approving the *S* for parallel behavior.

Probably crucial to the forward-moving procedure is the instantaneous detection of reflex voiding. Through the quick actions of the trainer in response to the signal some residual urine remains for deposit at the commode, so permitting a complete training trial. There are two aspects of control over bladder functions which deserve particular mention. When the focus of training is the voiding reflex occurring at locations outside the toilet facility, a child is required to learn to constrict the urinary sphincter or "turn off" the flow of urine, but he must also learn to relax the urinary sphincter and "turn on" the flow when he reaches the commode. The "turn off-turn on" control functions should be practiced in a sequence separated by only 3 or 4 seconds so as to be in concert with the trainer's immediate purpose—"don't do it here, do it over there."

These conditions for the practice of valve control are not present in the potting procedure. Under ordinary circumstances where little attention is paid to "turn off-turn on" behavior, valve control may not be adequate

for several years. Lovibond (1964, pp. 22–24) has discussed the research of the urologist Muellner (1960) which concerns the development of "voluntary" control of micturition (instrumental behavior). In this work Muellner, after completing more than a thousand fluoroscopic studies of maternally trained children, concluded that control was gradually acquired and did not represent the adult capacity for control until about age 6. Relevant to present considerations is the aspect of learning to initiate the flow of urine. Lovibond comments that there is "the common observation that young children will declare themselves unable to urinate while kept on a pot over long periods, only to void spontaneously almost immediately after being released." It is reasonable that the learning to initiate urination can be learned best when "involuntary" reflex voiding occurs in places other than the toilet facility so that the "turn off-turn on" behavior can be deliberately practiced in like form to that when the training will be completed.

A comparative summary of certain aspects of bladder training under a common potting procedure and the forward-moving procedure is as follows:

Potting procedure	*Forward-moving procedure*
1. Child is potted as a first training step and given approval for an event at the commode. Effort is to teach him to void—whether the bladder is full or not.	1. Child is given environmental training signals (discriminative stimuli) in association with a full bladder and quick movement to the toilet facility.
2. Voiding at inappropriate locations is capriciously detected, disapproved. High vigilance is required of the trainer.	2. Voiding at inappropriate locations is the focus of training —always detected, disapproved. Modest amount of vigilance is required of the trainer.
3. First learning is that of releasing urine for approval.	3. First learning is termination of urinary flow immediately followed by learning to release urine at the commode.
4. Elements of the learning task are practiced and acquired one at a time over a period of weeks or months. Trials are relatively inconsistent.	4. Elements of learning except those of clothing removal are acquired at once and all elements are practiced together within a few seconds on every trial.
5. Training trials are quite different from the behavior the child will exhibit when the training is complete.	5. Training trials closely approximate the final behavior expected of the child.

Procedure

Instrumentation. Auditory signal generators, worn by the child trainees, provided a signal which was coterminous with the flow of urine and was easily heard by both the trainer and the trainees.[1] This device was similar to one reported by Van Wagenen and Murdock (1966) except that the auditory signal precisely paralleled the flow of urine. A complex auditory signal, emitted only as urine continued to flow, originated from a transistor oscillator circuit feeding through a miniature speaker. The instrumentation was contained in a plastic box ($2 \times 3\frac{1}{8} \times 1\frac{1}{8}$ inches) attached to a belt which suspended a rubber or plastic urinal. The urinal containing sensing electrodes terminated in an orifice which allowed the urine to pass through and onto the floor or into the commode. The intensity of the signal was 90 db (re .002 dynes/cm^2) as measured across the speaker terminals. It was an easy competitor with the noises which children confined within a room usually produce.

Subjects. Four male and four female *Ss* ranging in age 4–9 years were selected from incontinent retarded children referred because personnel responsible for their care wanted them toilet trained. These children were reported to be most devoid of the desired habits among the children available. One additional male was referred by the Superior Court because his admission to the state institution for the retarded was contingent upon being toilet trained.

Subject characteristics. The degree of retardation in the *Ss* trained was generally profound. *S4*, a female, had some echoic speech and *S7* and 9 used three or four words. The remaining *Ss* had no speech. Two of the *Ss* had only recently learned to walk. Details are as follows:

S1—9 years, male, excellent walking, drinking, lowering and replacing of pants. No speech but response to command "stop!" Thwarting followed by *S* dropping to floor and repeatedly striking back against floor. No eye contact.

S2—7 years, male, shuffling walk recently acquired, drank small volume, always choked, always spilled. Could remove and return pants. Had severe cranial damage, was missing one ear and was partially paralyzed on one side. No speech but responded to commands "come here" and "stop."

S3—4 years, male, good motor capability. Held drinking cup without spilling, took large quantities of liquid. Did not remove or replace clothing. No speech but responded to name and command "come here."

S4—5 years, female, good motor capability. Held container and drank

[1] Mark 1 Toilet Trainer, Applied Psychology Associates, Box 983, Tempe, Arizona 85281.

large quantities of liquid without spilling. Lowered and replaced clothing. Imitative speech, responded to her name and commands "stop," and "come here."

S5—7 years, female, fair motor capability. Refused water, could hold cup but always dumped contents on floor. Could remove but not replace clothing. Frequent crying interfered with training. No speech, but responded to her name and commands "come here" and "no."

S6—8 years, male, fair motor capabilities. Drank large quantities of liquid with some spilling. Did not remove or replace clothing. No speech, but responded to his name and to command "come here."

S7—5 years, male, fair motor capability. Walk unstable, frequent falling. Held cup without spilling but drank little. Could lower pants but not return. Speech—a few isolated words. Responded to name and command "come here."

S8—5 years, female, fair motor capability. Unstable walk. Medication for seizures. Drank large quantities without spilling. Did not remove or replace clothing. No speech and no response to verbal commands.

S9—5 years, female, good motor capability. Drank moderate amounts without spilling. Lowered and replaced pants. Speech—few isolated words. Response to commands "stop" and "come here."

Subjects' entering behavior. Entering rates of urination were counted for a minimum of 5 days prior to the training trials. During this period *Ss* were clothed in cotton training briefs and upper garments only; placed three at a time in a room where they could move about freely, and observed daily for 3–4 hours; a period of time equal to that of the training to follow. Induced consumption of liquids was a part of baseline procedure as well as the training procedure in order to increase the frequency of urination.

Baseline urination events were recorded at one of six levels of response adequacy identical to those used for recording the results of training (see Table 1, The Training of *S*1). Level-1 represented the untrained state of urinating through the cotton brief and onto the floor. Level-6 was the criterion performance of autonomous toilet behavior; i.e., correct responding without prompts. Four intermediate steps represented approximations to the terminal behavior.

The nine subjects on whom measures were taken produced a baseline total of 311 level-1 events and six level-6 events with no intermediate behavior represented. Seven *Ss* performed entirely at level-1 during baseline observations while the remaining two *Ss*, hereafter designated *Ss* 1 and 2, produced four and two responses, respectively, at level-6, and 38 and 20 responses at level-1.

Training sessions and learning environment. Because economic considerations enter into the training of children, and because the subjects of this experiment were available in numbers, training was undertaken with three children at a time. The task of the experimenter-trainer was that of inducing the consumption of liquids by offering it frequently and waiting for the auditory signal to initiate a training trial. Trials occurred sufficiently often so that with the induced feeding of water, and the ordinary attentions children require, the experimenter-trainer was kept occupied with his task.

A small room adjoining a toilet facility was used for training. Toys were available and play activities were not generally restricted except at the occurrence of a signal which initiated a training trial. When one of the signal generators sounded, indicating that urine had begun to flow, the trainer made a rapid approach to the *S* from whom the sound originated, called loudly—"stop," grasped the *S* by the arm, walked him quickly to the commode, and placed the child in the sex-appropriate position for voiding. The procedure usually effected the termination of urine flow. Retention of some residual urine for deposit at the commode was a result of the combined moderate startle effects of the signal itself and of rapid exaggerated responses by the experimenter-trainer at the occurrence of the tone signal. Reinitiation of urination at the commode was facilitated by the *E* remaining quiet and by gentle stroking of the subject's back.

During early stages of training the signal device was worn continuously. This meant that urine flowing into the commode activated the signal just as it had done in the general training space. Each onset of the signal at the commode was a cue to the experimenter-trainer to verbally and gesturally approve (clapping the hands and other effusive gestures of approval common to the institution) the release of urine. When *Ss* had learned to urinate through the signal device into the commode and this behavior was reliably present, they were then trained to remove clothing immediately before taking the final stance.

Fluid consumption and the density of training trials. An unfortunate aspect of toilet training is the fact that the number of practice trials is limited by physiological processes closely correlated with fluid intake. At frequent intervals, therefore, *Ss* were offered as much water as they would drink and reinforced verbally and gesturally for drinking. Density of training trials was increased in this way. The frequency of baseline urinary events was 1.8 per subject per hour observed. As the training was undertaken, fluid consumption increased as did also the frequency of urinary events. During the later stages of training, it was not uncommon for *S* to have as many as 10 trials in a 3½-hour training session.

As a result the trainer had one urinary event to attend to approximately once every 10–15 minutes and used some of the intervening time to dispense drinking water.

RESULTS

*S*1 reached criterion performance of toileting behavior (level-6) in 19.5 training hours distributed over five sessions. Criterion performance included self-initiated walk in without prompts, controlled voiding from a standing position and management of clothing. At the beginning of the fifth session, the signal-generator-training-pants apparatus was replaced by standard cotton briefs. On this day there were eleven events at level-6 and one event at level-4. Among the nine trainees, this *S* demonstrated the most dramatic attainment of proper toilet habits as shown by the rapid upward sweep of the data from level-1 to level-6. Table 1 shows the data recorded for this subject and the progress of training. Each entry in the table refers to a urinary event. The entries are numbered separately for each day of training according to the order of their occurrence; i.e., 3, the third event of the day. Figure 1, *S*1 shows, in graphic form, the change in performance and its maintenance over 5 days.

The same data record as that obtained for *S*1 (Table 1) was employed also for *S*s 2 through 9. Liberty has been taken however, to convert the data from these tables to line graphs, disregarding the ordinal nature of the data. Figure 1, which presents the line graphs, is based upon plotted points obtained by averaging within levels-1 to 6, the learning trials of a single child for 1-day's performance.

Acquisition data for *S*2 as given in Figure 1 reveals that 13 sessions for a total of 44.5 hours were required to attain level-6 performance. The signal device was replaced by cotton briefs on session 13. In the post-training period *S*2 continued to manifest appropriate toilet behaviors in the institutional domicile through occasional accidents occurred with decreasing frequency. Follow-up 7 months after the last training session revealed no regression.

*S*3 achieved criterion behavior in 22 days. Progress, as shown in Figure 1, was uniform through level-4. At that point, and continuing for 13 days, *S*3 had difficulty learning to manipulate clothing and to successfully take the male stance at the commode. When this was accomplished, the behavior generalized to other environments readily and remained intact for 3 weeks after the last training session. The family left the community at that time and further follow-up was not possible.

Acquisition of toilet habits in *S*4 was without unusual incident. Level-6 was attained in 12 days and maintained for 2 months following the last session. Though this child was possessed of imitative speech (echolalia),

TABLE 1
THE TRAINING OF SUBJECT-1

Time—hours	1.41	5.00	4.75	4.33	3.33	2.75	4.25	4.50	4.08	4.50	1.58
Fluids consumed—ozs.	65	105	150	107	145	110	97	80	90	80	80
Response categories	Daily sequence of urinations under training conditions Day 1	2	3	4	5	6	7	8	9	10	11
6. Walk-in, remove clothes, urinate. No prompts.					2 3 . . . 12	3 5	1 . 2 . 3 . . 10 .	1 . 2 8 3 . .	1 . 2 . 3 . . 12 .	1 . 2 . 3 9 . .	1 2 3 4
5. Walk-in, remove clothes on command. Urinate in commode.						1 2 4					
4. Walk-in, urinate through urinal into commode.			5 13 6 7 10	2 7 3 . 5 . 6 13							
3. Gestural signal from *S*, then urinate on floor.			2		1						
2. Urinate, part on floor, part in commode.	1 2	2 3	1 9 3 11 4 12 8	1							
1. Urinate on floor.		1 4		4			6				

the speech did not prove to be exploitable in the training. Two months after training ceased, accidents began to occur which the parent attributed to inconsistent actions of a day-care babysitter.

*S*5 required the use of the signal generator for 15 days. Progress may well have been more rapid had it not been for a low density of training

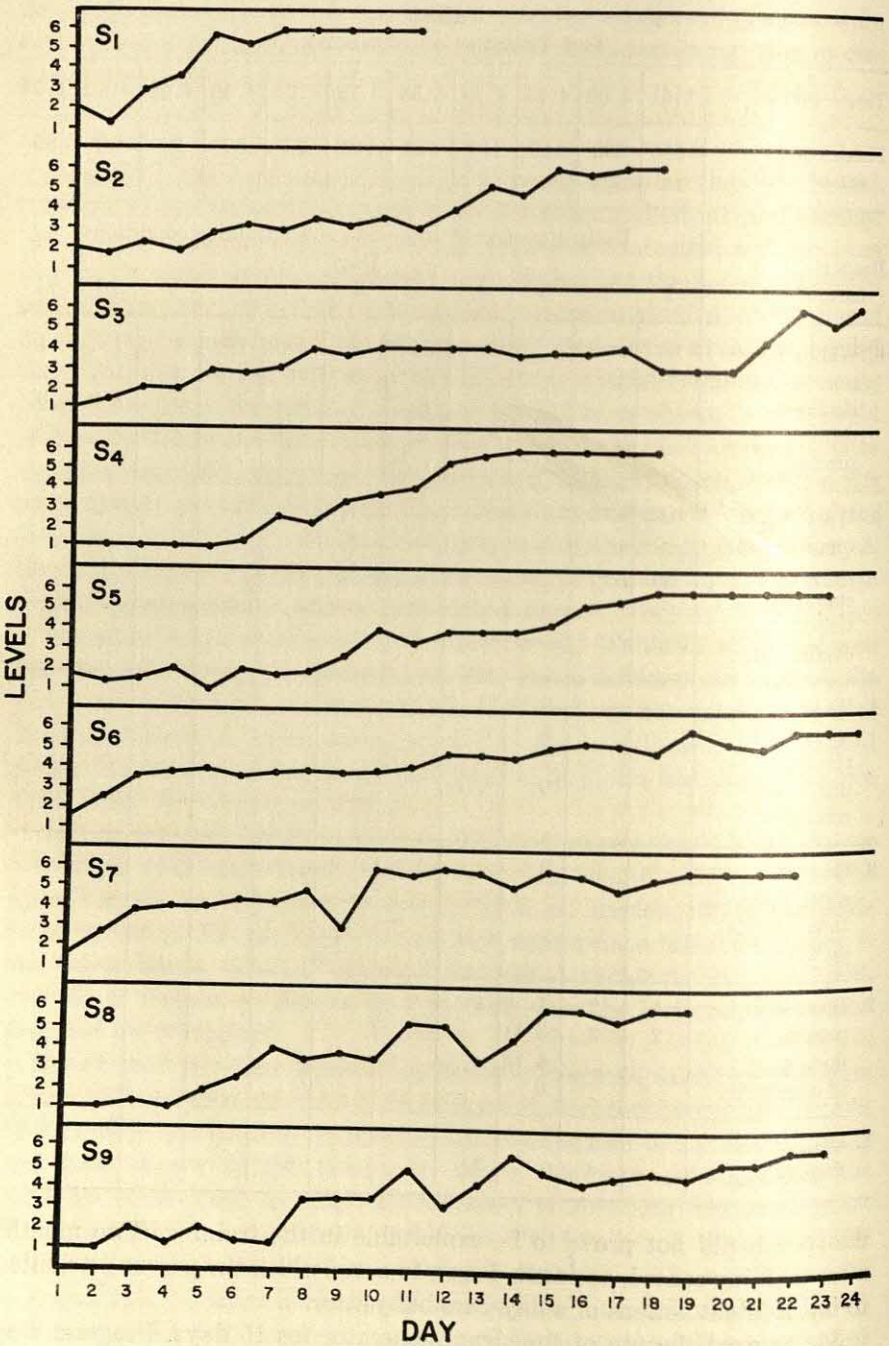


FIG. 1. The acquisition of toilet behaviors in nine mentally retarded children.

trials in consequence of low consumption of liquids. In the posttraining period of 5 months, appropriate toilet behaviors remained intact except for a 3-day period when accidents occurred and the parent reinstated portions of the training procedure for a few trials.

*S*6 was walking into the facility and completing urinations through the urinal into the commode after 3 days of training. Fifteen days were required to reach criterion behavior. Appropriate behavior was maintained in the posttraining period except that in the third month some inappropriate behavior recurred. During that period while the subject continued to manifest appropriate behavior, the mother began to prompt the child to go to the facility. This soon led to dependence on prompting. When the mother desisted, the problem was resolved.

*S*7 attained level-6 performance in 7 days. No unusual episodes occurred in the training and the behavior continued in the posttraining period. For a brief time a problem arose parallel to that with *S*6 when prompting from the mother disturbed the autonomy of the behavior. Again when the mother desisted, the problem was resolved.

Acquisition of level-6 behavior occurred for *S*8 on the ninth day and the behavior has been maintained to the present (one month) at which time some accidents are appearing.

Criterion was reached by *S*9 on the fifteenth day. A month following the last training session the behavior was intact.

DISCUSSION

Limitations on training without potting and without a signal generator. Systematic study of the performance of humans detecting minimal signals from visual displays reveals substantial incompatibility of the human and the task. Attentional lapses are common even when the vigil is limited to brief time periods (Buckner and McGrath, 1963). The visual sense is better suited to the presentation of complex messages when there is no time urgency in the transmission and the receiving human is free to maintain a visual posture. The aural channel, on the other hand, is particularly suited to simple messages requiring instantaneous detection and the suitability is still greater when the receiving human is unconfined or engaged in irrelevant activity, as mothers normally are or prefer to be (Chapanis, 1966).

In this light consider the immense history of toilet training as a domestic practice. Mothers everywhere have practiced it for centuries and the potting method which has survived has a substantial rational justi-

Six levels of response adequacy, identical to those in Table 1, are shown for each subject. A point on any one of the line graphs is an average of 1 day's performance for a single *S*.

fication. However, the nature of eliminatory processes and human ecology (use of clothing, stationary toilet facilities) being what they are, the mother has few options as to training method. Since the eliminatory event is generally concealed from rapid sensory detection, the mother does what these circumstances permit. She pots the child and keeps a visual vigil. The odds are against her accomplishing the task efficiently.

On several occasions, the senior author attempted toilet training by the "forward-moving procedure" without the use of the auditory signal generator; that is, by visual vigilance only. It is very difficult to do this, even for so short a period as 5 minutes, for the trainer must be closely attentive if he is to act quickly enough to complete a training trial. Several opportunities for training trials were missed in spite of firm intentions.

Maintenance of behavior. There is reason to believe that maintenance of appropriate toileting behavior is dependent upon the degree to which parents and attendants expect the child to manifest the behavior and intervene only if the child fails to perform as expected. If caretakers revert to earlier prompting or potting procedures, the behavior breaks down. Toileting behavior is dependent also upon having free access to a toilet. *S*1, for example, was resident in an institution where he was confined within the locked doors of a ward for the most disturbed children in the institution. Since that ward had no toilet facility freely accessible to its residents, *S*1 could not manifest the behavior of which he was capable. Demonstration to institutional personnel that the child was capable of toileting self-care was regarded as an unexpected curiosity but had no effect on the practice of diapering all children in that ward, including *S*1. In one case, *S* residing at home, the parent kept the bathroom door locked because the child played with the objects found there. The very adequate toileting behavior obtained in training sessions was not reinstated in the home until the bathroom was cleared of extraneous objects and the door could be left open. The training of parents and attendants to maintain the behavior taught in the training sessions is a separate problem which requires additional investigation.

Reinforcement and information in toilet training. Diurnal toilet training requires the learner to perform a complex sequence of behaviors. In order to fulfill what is expected of him the learner requires considerable information. It is on this point that the potting procedure for toilet training has a severe limitation. Common practice frequently calls for reinforcing children with approval and sometimes with sweets when, after a wait at the commode, they eventually void. Even though the reinforcer is immediately contingent, it provides only a limited amount of information to the learner. He may be informed that the terminal voiding act

is approved but such reinforcement is little help in approving those required responses that are preliminary to the terminal voiding response such as restraint of reflex voiding, going without prompts to the toilet facility when the need arises, and removing and replacing clothing without soiling. Similarly, changes in quantity or quality of reinforcement are unlikely to benefit the learning. What the learner requires is quick practice of the entire behavior sequence. From the rapid performance and consequent social approval sufficient information is provided to the learner to allow rapid formation of discriminations. Social approval, sometimes regarded as a relatively weak reinforcer, proved in this study to be quite adequate and no necessity arose for the use of consumables as reinforcers. Reconceptualizing the task in terms of the informational content which may be available under potting conditions and under the training conditions described here is a probable basis on which to account for the rapid learning of toilet habits in this study.

REFERENCES

Buckner, D. N., and McGrath, J. J. *Vigilance: A symposium.* New York: McGraw-Hill, 1963.

Cassell, R. H., and Colwell, C. N. Teaching the profoundly retarded self-help activities by behavior shaping techniques. In Gerald J. Bensberg (Ed.), *Teaching the mentally retarded.* Atlanta: Southern Regional Education Board, 1965.

Chapanis, A. *Man-machine engineering.* Belmont, Calif.: Wadsworth, 1965.

Dayan, M. Toilet training retarded children in a state residential institution. *Mental Retardation,* 1964, **2,** 116–117.

Ellis, N. R. Toilet training the severely defective patient: An S-R reinforcement analysis. *American Journal of Mental Deficiency,* 1963, **68,** 98–103.

Giles, D. K., and Wolf, M. M. Toilet training institutionalized, severe retardates: An application of operant behavior modification techniques. *American Journal of Mental Deficiency,* 1966, **70,** 766–780.

Hunziak, M., Maurer, Ruth A., and Watson, L. S., Jr. Operant conditioning in toilet training of severely retarded boys. *American Journal of Mental Deficiency,* 1965, **70,** 120–124.

Lovibond, S. H. *Conditioning and enuresis.* New York: Macmillan, 1964.

McGraw, Myrtle B. Neural maturation as exemplified in achievement of bladder control. *Journal of Pediatrics,* 1940, **16,** 580–590.

Muellner, Richard S. Development of urinary control in children. *Journal of the American Medical Association,* 1960, **172,** 1256–1261.

Pumroy, D. C., and Pumroy, Shirley S. Systematic observation and reinforcement technique in toilet training. *Psychological Reports,* 1965, **16,** 467–471.

Spock, B. Toilet training after eighteen months. *Redbook,* July 1968, 22–27.

Van Wagenen, R. Keith, and Murdock, Everett E. A transistorized signal-package for toilet training of infants. *Journal of Experimental Child Psychology,* 1966, **3,** 312–314.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 160–172 (1969)

A Developmental Study of Hypothesis Behavior and Focusing[1]

PETER D. EIMAS
Brown University

Children from Grades 2, 4, 6, and 8 and introductory psychology students were administered eight 16-trial hypothesis testing problems. The stimuli and procedure closely followed those used by Levine (1966). The major findings were as follows: (a) *Ss* from all grade levels formulated and used hypotheses reliably; (b) the extent to which hypothesis behavior was manifested increased significantly with developmental level; (c) the ability to predict outcome-trial responses from the blank-trial responses was also significantly related to development; and (d) focusing or information processing ability was differentiated by age, although only for the blank-trial sequences after the second and third outcome trials. Further analyses of the age-related difference in the focusing functions revealed that the general effect due to development could be explained by age-related differences in such processes as coding, recoding, and the retention of stored outcome information.

Levine (1963, 1966) has presented a methodology for the study of hypothesis behavior and focusing within the context of discrimination learning that provides simple and direct measures of both forms of behavior. However, in order to facilitate the understanding and explication of the procedure, a brief description of the rationale or model underlying Levine's methodology will be presented prior to the methodology itself. Levine has assumed that in a discrimination learning situation, *S* will select some attribute or feature of a multidimensional stimulus array and predict that this is the correct or relevant feature and respond accordingly. This selected attribute is what Levine has defined as an hypothesis (*H*). Levine has further assumed it to be one form of mediational response that can control or guide instrumental responses of choice. By instructions and pretraining the set of *H*s from which *S*

[1] This investigation was supported by Grant HD 03045 from the National Institute of Child Health and Human Development, U. S. Public Health Service. The author would like to express his appreciation to Mr. Kenneth L. Mackenzie, Superintendent of Schools, Highland Park, New Jersey, for his generous support and cooperation and to Mrs. Joyce Weinberg and Miss Suzanne Henry for their assistance in the collection of the data.

samples is made finite and known to *E*. Measurement of a particular *H* without the confounding effects of reinforcement and nonreinforcement is possible, since *S*s had been found to maintain the same *H* across consecutive nonoutcome ("blank") trials. This phenomenon has been termed the "Blank Trials Law" (Levine, Leitenberg, and Richter, 1964). In other words, after no outcome *S* behaves as if *E* had said *S* were correct in that he retains and uses the same *H* again.

As for actual procedure, Levine (1966) presented his *S*s with a series of 16-trial problems in which sets of four blank trials were bounded by outcome trials.[2] The stimuli were four-dimensional patterns with two levels or values along each dimension. With four-dimensional bivariate stimuli there are eight different stimulus arrays possible, four of which are shown in Figure 1. The remaining four are constructed by reversing the left-right positioning of the two stimulus patterns within each card. Each set of four stimuli is internally orthogonal: Each level of every dimension is paired exactly twice with each level of every other dimension.

A typical problem of 16 trials uses one set of four arrays for the out-

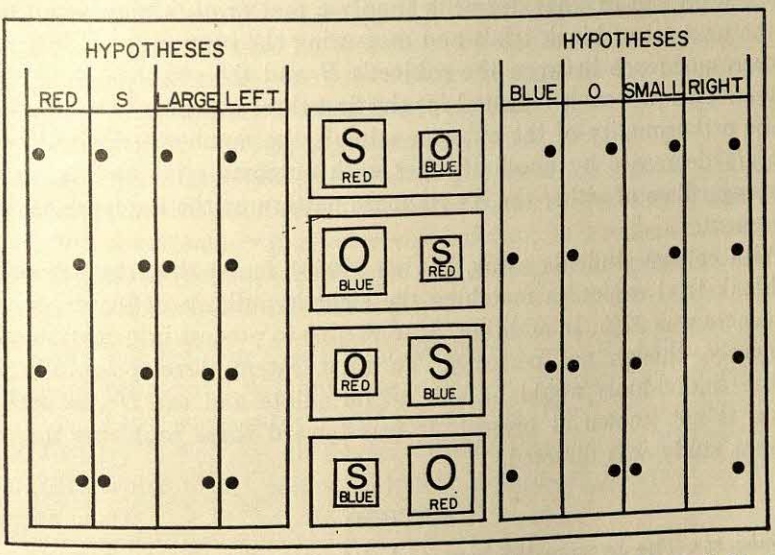


FIG. 1. The eight response sequences defining the eight simple *H*s available to *S* when the four stimulus arrangements are administered consecutively and without outcome information.

[2] Trial 16 is technically not an outcome trial, since an outcome was not always given. On half of the trials, the *S* was told he was correct and on the remaining half nothing was said; i.e., it was a blank trial. However, since this trial used stimuli from the same set as the other outcome trials and since positive outcomes were sometimes given, it will be referred to as an outcome trial.

come trials, Trials 1, 6, 11, and 16, and the second set for the blank-trial sequences, Trials 2–5, 7–10, and 12–15. If we assume for purposes of explication that the stimuli shown in Figure 1 represent a set of four blank trials, then the subject's hypotheses may be shown to be easily measured. Any pattern of responding matching one of the eight patterns shown in Figure 1 by definition corresponds to one of the eight simple *H*s available. For example, a sequence of left, right, left, right would define the *H* red. Response patterns not conforming to those shown, any pattern of three lefts and one right or the converse, are assumed to occur through error. Thus, by analyzing the response sequences on blank trials, evaluation of the extent to which *S*s employ *H*s can be made.

As for *S*'s efficiency in focusing or processing information, this can be assessed in the following manner. Assume now that the stimuli of Figure 1 represent outcome trials. If *S* chose the left stimulus of the first pair on Trial 1 and was informed that he was incorrect, then four *H*s would be logically correct and four incorrect, the correct four being right, small, blue, and O. To determine whether the subject has utilized the outcome information and to what degree is simply a matter of determining the *H* on the next set of blank trials and measuring the proportion of instances of correspondence between the subject's *H* and the set that is logically correct. This process is repeated for the first three outcome trials. Because of the orthogonality of the outcome stimuli, the number of logically correct *H*s decreases by one-half after each outcome trial, this being the case regardless of either the *S*'s response pattern or the sequence of reinforcement.

With college students as *S*s, Levine (1966) found that the proportion of blank-trial responses matching the eight hypothesis-defining response sequences was 92%. In addition, *S*s were able to process information quite efficiently, though not perfectly. To what extent developmentally less mature individuals would be able to formulate and use *H*s, as well as focus, is not known at present. It was toward these problems that the present study was directed.

METHOD

Subjects. The *S*s were 128 boys and girls attending Grades 2, 4, 6, and 8 in the public schools of Highland Park, New Jersey. Thirty-two *S*s were drawn from each grade. A summary of their characteristics is shown in Table 1. In addition, 32 girls from the introductory psychology course at Douglass College were tested.

Stimulus materials. The experiment consisted of a series of 12 problems —four practice and eight experimental. All problems were four dimensional with two cue values along each dimension. The dimensions were

TABLE 1
SAMPLE CHARACTERISTICS

	CA (months)		MA (months)		IQ	
Grade	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
2	93	5.08	98	10.30	105	11.36
4	118	6.33	123	13.70	105	12.80
6	139	6.16	152	17.30	110	14.69
8	162	5.44	185	18.17	113	11.24

color, letter, size, and position. Twelve different colors and letters were used. Each was used twice in different randomly determined combinations. The stimuli were constructed by cutting 1- or 1/2-inch squares from Bourges Cutocolor sheets and attaching them to 3×5 cards. Onto the squares were attached black Letraset letters either 3/4- or 3/8-inch in height.

Procedure. Each *S* was tested individually. In order to insure that *S*'s set of potential *H*s conformed to the eight simple *H*s shown in Figure 1, detailed instructions, similar to those used by Levine (1966), were given. The only major differences were an attempt to simplify the language for the younger children and to emphasize repeatedly that ". . . one of the two colors, or letters, or sizes, or positions is always correct."

After the instructions, the four practice problems were given. The first two problems consisted entirely of outcome trials. In Problem 1, size was the relevant dimension, while color was relevant in Problem 2. On the third and fourth practice problems, the blank-trials procedure was introduced with the following instructions:

> "In the last problem I said right or wrong after each card. In the remaining problems I will not always tell you whether you are right or wrong. After some cards I will say nothing. Don't let this bother you. Just try to be right all the time. Remember you are still to try and learn which of the colors, or letters, or sizes, or positions is the correct one."

The relevant dimensions were letters and position in Problems 3 and 4, respectively. The *S*s were trained to a criterion of six consecutive correct responses on each practice problem. If an *S* failed to attain criterion within 16 trials, hints were provided; e.g., "Why don't you try one of the colors?" In this manner all *S*s received practice in using each of the dimensions and in using simple attributes for solution.

After completion of the practice problems, the eight experimental problems were begun, each being preceded by a reminder that *E* would

not always tell *S* whether he was correct or not and that it was one of the colors, etc., that was correct. Each experimental problem had exactly 16 trials, and consisted of four outcome trials, occurring on Trials 1, 6, 11, and 16, and 12 blank trials, arranged in sets of four and presented on Trials 2–5, 7–10, and 12–15. Each set of blank trials had a different random order. One set of internally orthogonal stimuli was used for the outcome trials and the other for the blank trials, thereby insuring that specific stimulus arrangements used on the blank trials were never reinforced.

The *E* said right or wrong on the first three outcome trials according to a prepared schedule. Each of the eight possible orders of right and wrong was used once per subject. The eight orders were assigned to eight different problems for each eight *S*s forming 8×8 Latin squares. The *E*'s response to the last outcome trial was either correct or blank, each occurring randomly and equally often.

On all trials, *S*s were required to point to one of the two stimulus patterns on each card that they believed to be correct. At the conclusion of the experiment, the children were rewarded with a pen or package of Charms.

RESULTS AND DISCUSSION

Blank-trial data. There are two classes of response sequences from the sets of blank trials: One corresponds to the eight simple *H*s and the other corresponds to the eight, 3–1 patterns of lefts and rights not defining a simple *H*. Each *S* received 24 blank-trial sets. The mean percentages of response sequences conforming to the simple *H*s were 71, 73, 77, 79, and 88% in order of increasing developmental level. Each of these scores differs significantly from 50%; i.e., the expected score had *S*s been randomly generating responses rather than responding systematically in terms of *H*s ($t > 9.00$, $p < .001$, in each instance). A single-classification analysis of variance performed on number of *H*s revealed a significant effect due to age at better than the .001 level ($F = 8.4$, $df = 4{,}155$). Individual comparisons among groups revealed no reliable differences between Grades 2, 4, and 6. The eighth-grade children formed significantly more *H*s than did the second and fourth graders. The college students performed reliably better than all other groups. Furthermore, the college students of the present study formulated *H*s at very nearly the same rate as did the *S*s in Levine's (1966) original study, 88 vs. 92%, respectively.

Levine has assumed that non-*H* response patterns are a result of error in attempting to formulate *H*s, and not a result of random responses by *S*s. He further assumed that there will always be some small probability

of error on any trial. Given that the college students had 12% of their response patterns in the 3–1 category, the error rate per trial is then 3% (that is, each response sequence involves four trials of which only one is assumed to be an error). The error rate per trial increases from this figure to a maximum of 7% for the second graders.

Another measure of the extent to which *H*s determine choice responses is the *E*'s ability to predict outcome-trial responses on the basis of the blank-trial data. For example, if an *S* showed a pattern of responses defining the *H* red, then on the succeeding outcome trial one would expect a choice response to the stimulus pattern containing red. Selection of the red stimulus in this instance defines a correctly predicted outcome trial. The mean percentages of predicted outcome trials were 74, 67, 78, 88, and 95% in order of increasing development. Once more each of these scores differs from the chance expectation of 50% ($t > 3.8$, $p < .001$, in each instance). A single-classification analysis of variance was performed on the percentage of correctly predicted-outcome trials normalized by the arc-sine transformation. The developmental effect was highly reliable ($F = 12.9$, $df = 4,155$, $p < .001$). Individual comparisons revealed all comparisons to be significant except those between the second and fourth and the second and sixth grades.

Comparison of the blank-trial data with the predicted-outcome scores reveals an inconsistency. The erroneously predicted outcome trials can be considered an estimate of the error rate per trial. As is apparent, this estimate for the three younger groups (about 25%) is considerably higher than the error rate obtained from the blank-trial data (about 6–7% for the same age groups). The differences in error estimates for the eighth graders and college students are smaller, being approximately 7 and 1%, respectively. To reconcile this inconsistency, the entire series of blank trials could be treated as a whole and a non-*H* pattern treated as a single error. This reasoning reduces the difference between error estimates to about 5%. Alternatively, it could be assumed that developmentally less mature *S*s may at times tend to forget from trial to trial the *H* previously used. Evidence supporting a memory deficiency in younger children will be discussed later. It is of interest to note, however, that this assumption can account in part for two other findings that were at variance with Levine's (1966) data, namely, the effects of outcomes on the retention and elimination of *H*s.

Levine found that a positive outcome resulted in *S*s repeating the previously tested *H* with a probability of .95, while after a negative outcome trial the probability of repeating the preceding *H* was only .02. In the present study, after a positive outcome trial the college students tended to repeat the previous *H*, the probability of this event being

slightly better than .90. However, this probability decreased significantly with decreasing age, reaching about .60 for second graders ($F = 7$ $df = 4,155$, $p < .01$). After a negative outcome trial, the probability a repeated H was .08 for college students, .15 for fourth, sixth, and eight graders, and .18 for second graders. The differences were not significant however, in as much as the majority of the 160 *Ss* either never repeated an H after a negative outcome (85 *Ss*) or did so only once (51 *Ss*).

The failure of the younger *Ss* to replicate the earlier findings of Levine with regard to these measures does not preclude further analyses of more complex functioning; i.e., focusing or maximally efficient information processing. However, this failure does demand that explanations of any obtained discrepancies from an ideal information processing model also account for the differences in H behavior after outcome trials.

Focusing. The measure of focusing or information processing is essentially a measure of the extent to which *Ss* are able to utilize the outcome information on the blank trials. As was explained above, after the first outcome trial, four *Hs* are defined as logically correct, regardless of either *S*'s response or the outcome. Two *Hs* are correct after the second outcome trial, and one H after the third outcome trial. If one finds the proportion of instances in which a subject's H corresponds to one of the set of logically correct *Hs*, it is then possible to determine the actual size of the hypothesis set from which the subject is sampling. The equation is simple: The proportion of correct *Hs* equals the number of logically correct *Hs* (four, two, or one) divided by the actual size of the hypothesis set from which the subject samples. The latter term is the single unknown in the equation.

In computing the proportions of logically correct *Hs*, we have restricted our sample to those *Hs* occurring after negative outcome trials. The reasons for this decision are twofold. First, after a positive outcome *Ss* do tend to repeat the same H and this may artificially improve the measure of information processing. Second, by using *Hs* only after negative outcomes a test of two ideal or limiting models of information processesing is possible: Restle's strategy selection model (Restle, 1962) and the information processing model outlined above, which is very similar to the focusing strategy described by Bruner, Goodnow, and Austin (1956). The Restle model makes the assumption that *S* samples one strategy (hypothesis) from a set of strategies (hypotheses) and after a wrong response; i.e., negative outcome, resamples but only after returning the incorrect strategy to the set of strategies. The Restle model thus leads to the prediction of no systematic reduction in the size of the set of hypotheses after negative outcome trials.

The first panel of Figure 2 shows the predicted functions for the two

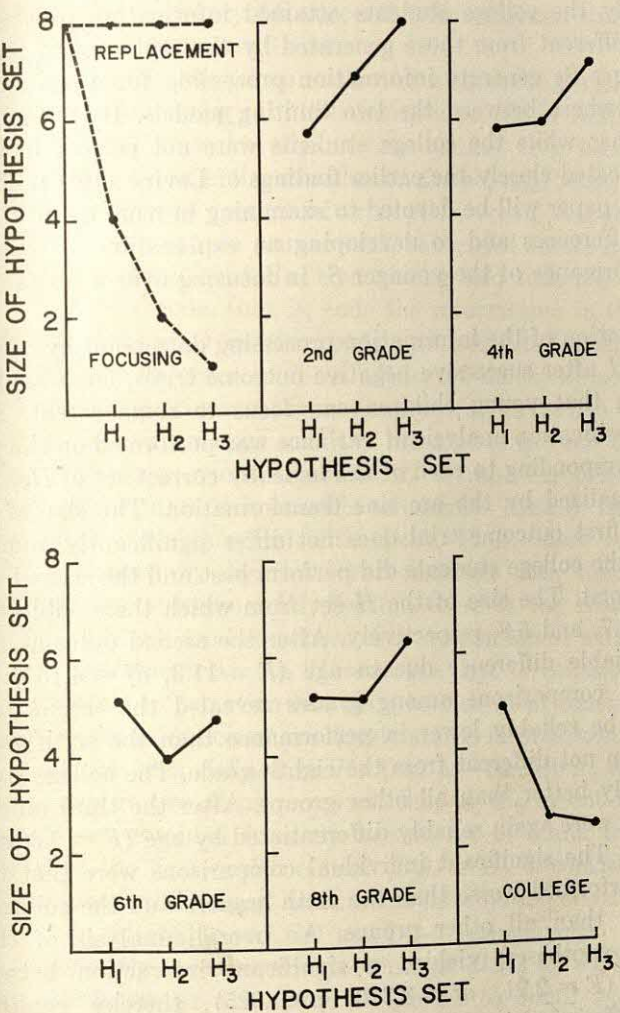


FIG. 2. The size of the H set from which Ss are sampling after a negative outcome trial as a function of age and the blank-trial sequence H_1, H_2, and H_3.

models—replacement and focusing. The remaining panels depict the size of the H set from which each H is sampled after negative outcome trials as a function of grade level. As is apparent, none of the functions matches the replacement model at all points, and only the college population generates a function approximately the focusing strategy across all three Hs. Statistical analyses revealed that the replacement model may be rejected as describing the information processing strategy for all Ss at H_1 and for the college students, sixth, and eighth graders at H_2. However,

at H_3 only the college students attained information processing scores reliably different from those generated by the replacement model. Thus, the younger *Ss* generate information processing functions for the most part somewhere between the two limiting models. It is also of interest to note that while the college students were not perfect focusers, their data replicated closely the earlier findings of Levine (1966). The remainder of the paper will be devoted to examining in more detail the developmental differences and to developing an explanation for the relatively poor performance of the younger *Ss* in focusing over a series of outcome trials.

Examination of the information processing data point by point, that is, for each H after successive negative outcome trials, lends support to the contention that young children can focus to some extent. A separate single-classification analysis of variance was performed on the percentage of *Hs* corresponding to each of the logically correct set of *Hs*. The scores were normalized by the arc-sine transformation. The size of the H set after the first outcome trial does not differ significantly among groups, although the college students did perform best and the second and fourth graders worst. The size of the H set from which these subjects sampled was 4.9, 5.7, and 5.8, respectively. After the second outcome trial, there was a reliable difference due to age ($F = 11.3$, $df = 4{,}155$, $p < .001$). Individual comparisons among grades revealed the second and fourth grades to be reliably lower in performance than the sixth grade which was in turn not different from the eighth grade. The college students did significantly better than all other groups. After the third outcome trial, the results were again reliably differentiated by age ($F = 9.3$, $df = 4{,}155$, $p < .001$). The significant individual comparisons were that the second graders performed worse than the sixth graders and the college students did better than all other groups. An overall analysis of these data, groups $\times$ hypotheses, yielded a significant interaction between these variables ($F = 2.21$, $df = 8{,}310$, $p < .025$), thereby confirming the marked differences in the shape of the functions with age that are apparent from inspection of Figure 2.

The focusing data indicate that *Ss*, while entertaining a single H, often apply information concerning its correctness not only to that H but also to other attributes (*Hs*) correlated with the tested H. The extent to which *Ss* are able to do this, however, is limited by developmental level.

The problem now remains to explain the considerable differences in the focusing functions associated with developmental level. It is possible to appeal to a stage-like progression in focusing ability, using as evidence the marked discontinuity in the shape of the obtained focusing functions. However, an appeal of this nature would be more descriptive than explanatory. A more meaningful approach would be to explicate the

specific processes necessary for successful focusing and to evaluate them with respect to developmental level.

Levine (1966) has offered an outline of a theoretical model describing the processes or psychological mechanisms involved in focusing. The discussion below will draw heavily on this outline. It is assumed that on an outcome trial, *S* codes the stimulus of choice; e.g., large, red, S on the left, by means of some perceptual or implicit verbal response. On being informed that his response was correct, the coded information then becomes the set of functional *H*s. Upon being told that the response was incorrect, *S* must recode, that is, code the information in the unselected stimulus (small, blue, O on the right) and use this information as the functional *H* set. In order to focus efficiently across outcome trials, at least two additional processes are necessary: (a) Retention of previously coded or recoded information and (b) the ability to code (recode) only the intersect of the previous functional *H* set and the present stimulus.

With the present data it is possible to examine some of these processes for evidence of developmental differences. First, there is the coding process as measured by *H*s that are consistent with the immediately preceding positive outcome trial. Second, there is the recoding process as measured by blank-trial responses that are consistent with the preceding negative outcome trial. An *H* response that is consistent with the preceding outcome information has been termed local consistency (LC). Third, there is the retention of previously stored information. For present purposes this has been defined as *H*s that are consistent with outcome information one or two steps removed; that is, the extent to which the functional *H* set obtained from the second set of blank trials (H_2) is consistent with the first outcome trial (O_1) or H_3 is consistent with O_2 or O_1. Consistency with respect to outcome information one and two steps removed will be referred to as LC-1 and LC-2, respectively.

Figure 3 shows the mean correct percentages of *H*s for each of the local consistency measures as a function of the set of blank trials and age. The difference in ability to utilize positive as opposed to negative outcome information was reliable for both group and individual comparisons. For example, of 30 possible group comparisons (six within each of five developmental levels), 29 showed a difference of at least 5% favoring positive outcome information ($p < .001$) and of these, 15 showed a difference of greater than 10%. Moreover, there was no systematic change in the magnitude of the differences with age; the mean differences were 9.2, 12.0, 10.3, 15.2, and 11.7% in order of increasing age. With respect to individual performances, 125 of 160 *S*s performed at a higher level after positive outcome trials ($p < .001$) with no reliable differences in frequency as a function of age.

For all three LC measures, analyses of variance, performed on the

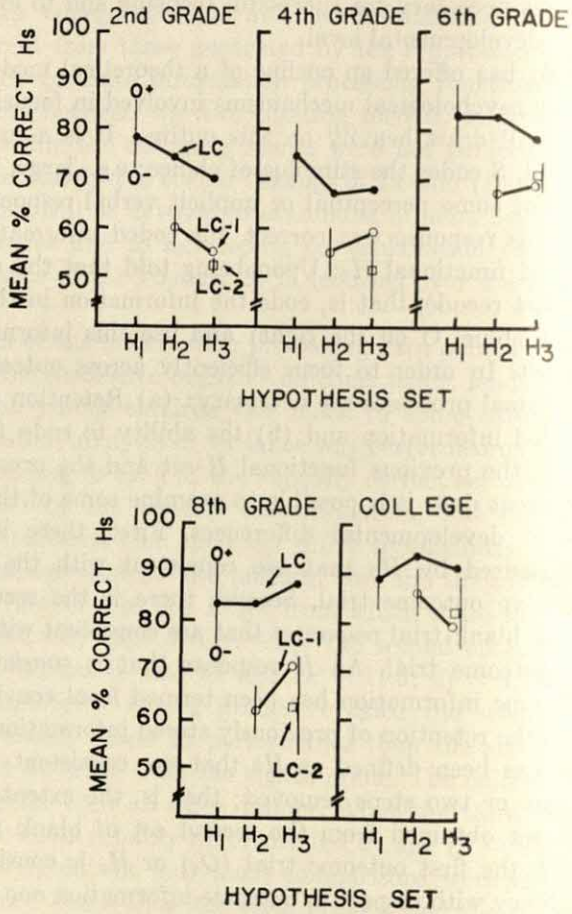


FIG. 3. Mean percentage *Hs* consistent with the immediately preceding outcome trial (LC), the outcome trial one step removed (LC-1) and two steps removed (LC-2). H_1, H_2, and H_3 represent the *Hs* given to the first, second, and third sets of blank trials. The vertical lines drawn at the initial point of each function show the mean differences in performance between positive (O+) and negative outcome trials (O—).

transformed percentage scores, revealed reliable improvements in performance with increasing age ($F > 9.0$, $df = 4,155$, $p < .001$, in all instances). In general, individual comparisons revealed the following: (a) The second and fourth grades did not differ reliably, but both groups did significantly poorer than all other age levels; (b) Grades 6 and 8 were not significantly different from each other but did differ significantly from the college students. In the LC-1 condition there was some indication of a change in slope with changes in developmental level ($F = 2.8$, $df =$

4,155, $p < .05$). However, the slope variations were not systematically related to age and interpretations are, as a consequence, difficult. In summary, the LC data analysis indicated that the ability to code and recode information with reference to the immediately preceding outcome trial increases with developmental level. Analyses of the LC-1 and LC-2 data indicate that retention of previously stored information is likewise a positive function of developmental level.

Further evidence of retention effects may be seen in Figure 3 in the overall decrement in performance with the remoteness of the outcome information. Statistical confirmation of this observation was obtained by an analysis of variance performed on the H_3 data, normalized by the arcsine transformation. The effects of both age and degree of remoteness were reliable at better than the .001 level ($F = 21.1$, $df = 4,155$, and $F = 20.0$, $df = 2,310$, respectively). Individual comparisons revealed the LC condition to be reliably better than either the LC-1 or LC-2 conditions, with the latter two groups not differing significantly. With regard to individual grade differences, there were three nonsignificant comparisons, those between Grades 2 and 4, 6 and 8, and between Grade 8 and the college students. However, the latter comparison just failed to attain significance. The interaction of these variables, age and remoteness or LC condition, failed to approach statistical reliability.

The analysis of focusing behavior into component processes has revealed that the relatively poor performance on a complex task by developmentally less mature subjects may have been substantially the result of a summation of deficiences of degree in component processes. That is, the rather orderly decline in focusing with decreasing age was mirrored almost perfectly by an orderly decline in the individual processes of coding, recoding, and retention of coded information. Indeed, were it possible to measure each of the component processes, it does not seem unreasonable that a quantitative account closely approximating the changes in focusing behavior with age would be available. Moreover, this manner of approach to complex behavior, similar in principle to that offered by Gagne (1968) for conservation problems, provides a more detailed and complete explanation than do models that posit the relatively sudden appearance of new and qualitatively different abilities.

REFERENCES

Bruner, J. S., Goodnow, J. J., and Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Gagne, R. M. Contributions of learning to human development. *Psychological Review,* 1968, **75,** 177–191.

Levine, M. Mediating processes in humans at the outset of discrimination learning. *Psychological Review,* 1963, **70,** 254–276.

LEVINE, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 1966, **71**, 331–338.

LEVINE, M., LEITENBERG, H., AND RICHTER, M. The blank trials law: The equivalence of positive reinforcement and nonreinforcement. *Psychological Review*, 1964, **71**, 94–103.

RESTLE, F. The selection of strategies in cue learning. *Psychological Review*, 1962, **69**, 329–343.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 173–179 (1969)

Stimulus Complexity as a Determinant of Visual Attention in Infants[1]

ADRIENNE MOFFETT[2]
McMaster University

The length of time that infants fixated patterns of varying complexity was studied in a series of experiments using stimuli made up of black lines on a white ground. In Exp. I, the effects of the number of horizontal lines and the regularity of their arrangement was examined. In Exp. II, the arrangement of the lines was varied while keeping the number of resulting white rectangular spaces constant. Experiment III served as a control for the introduction of vertical lines in the patterns of Exp. II. In Exp. IV, the number of white rectangular spaces was varied while keeping the number of lines constant. The results of these studies showed that infants looked longer at patterns with a greater number of lines, regardless of whether they were arranged regularly or irregularly, and the number of white rectangular spaces created by the lines was a particularly important determinant of visual attention.

Many attempts have been made to determine visual preferences in infants, but there have been problems in isolating the critical stimulus properties. Several studies have been concerned with presenting stimuli which vary in shape, (Fantz, 1958, 1961; and Spears, 1964), and the complexity of the patterns has been suggested as an important parameter which attracts attention in infants (Berlyne, 1958; Fantz, 1966; and Hershenson, 1964). There has not always been agreement in defining the complexity of the stimuli, but from direct experimentation and from *post hoc* interpretation of the above experiments on shape it appears that infants may pay more attention to those stimuli made up of a larger number of separate "parts."

Some recent work concerned with isolating stimulus properties has employed modifications of an apparatus used by Staples (1932) in which the stimuli are presented in pairs, and the infants' preferences are determined either by the direction of the first glance (Berlyne, 1958) or by the

[1] This study is based upon a thesis submitted to McMaster University in partial fulfilment of the requirements for the Master of Arts degree. The writer expresses her appreciation to Dr. E. Wardwell-Ames for her helpful guidance and to G. Saayman for help in the testing.

[2] Now at the Institute of Psychiatry, De Crespigny Park, Denmark Hill, London, S.E. 5, England.

length of fixation (Fantz, 1958, 1961; Hershenson, 1964). In the present study the length of the visual fixation is used to determine the infants' preferences, but the stimuli, although shown two at a time, are presented in such a way that one of each pair is the patterned stimulus and the other is a control blank. This has the advantage of giving an absolute measure of preference for each stimulus.

In this study, complexity has been varied in terms of the number of parts; that is, number of lines and number of white rectangular spaces created by these lines, in four separate experiments.

METHOD

Subjects. Physically normal infant *Ss* were volunteered by their parents who were paid for participation in the experiment. Age was the criterion used for choosing the *Ss*: Their names were obtained from the Birth Announcements column of the local newspaper. Those *S*'s who failed to complete the experimental series because of crying or falling asleep were disqualified. Different groups of *Ss* were used in the inter-observer reliability study and in the four experiments to be described. All involved the same apparatus and general procedure.

There were 10 *Ss* in each experiment. The age range at the time of testing was 10–14 weeks in Exp. I; 9–19 weeks in Exp. II; 11–15 weeks in Exp. III; and 10–17 weeks in Exp. IV.

Apparatus and general procedure. The apparatus was a modified version of the "baby box" used by Fantz, Ordy, and Udelf, 1962. Details of this and the general procedure are described by Saayman, Wardwell-Ames, and Moffett, 1964.

All infants were placed face-up in the crib by their mothers. The crib was then rolled under the chamber to a position where the infant's head was directly below the point at which the stimulus patterns were to appear. The patterns were not visible to the infants at this stage because they were covered by a blind. When the observer had noticed, through a peep hole in the lid of the chamber, that the infant had looked up at the center hole in the blind, this latter was released and two stimuli each 6 inch square were exposed, thus beginning a trial of 30-second duration. In each trial one stimulus was a blank white card and the other was the stimulus pattern composed of 1/8-inch wide straight black lines arranged in different ways depending on the experiment. Each was paired twice, once on the left and once on the right with the blank in all experiments. The stimulus patterns were never directly compared with each other.

The patterns reflected on the cornea of the infant's eye. When the eye was directed towards a particular stimulus, the image of that pattern

overlapped the pupil as seen through the peep hole. This overlap of image and pupil was the criterion of fixation. An observer recorded the amount of time that each stimulus was fixated. The blind was drawn back at the end of the 30-second trial. There were two *E*s who worked independently throughout, one inserted the stimuli and the other recorded the length of visual fixation.

The amount of time spent looking at each of the stimulus patterns was converted to a percentage of the total time spent looking at that stimulus and at the control blank. An analysis of variance was performed on these percentages (Winer 1962).

The details of the patterns used in the various experiments will be described under the Results; however, the stimulus patterns are shown in Figs. 1–3.

A previous separate study had been performed to test for inter-observer reliability using 10 *S*s ranging in age from 11–15 weeks. The apparatus was the same as in the other experiments except that the center portion of the chamber ceiling was replaced by a window which allowed two *E*s to record independently the length of time that the infants fixated patterns consisting of formalized faces (Fantz, 1961). An agreement score was calculated by counting the number of seconds per 30-second trial that the two observers agreed. The scores were totaled and converted to a percentage of the total stimulus presentation time. The percentage scores showed that the two observers agreed 94% of the time during 60 trials of 30-second duration.

Stimulus material. Experiment I was designed to investigate the effects of two variables, number of parts, and regularity of arrangement of these parts on visual attention in infants. There were four stimulus patterns (see Fig. 1). Two of these were composed of 5 lines and the other two of 11 lines. Within each set of two, one contained the specified number of horizontal lines arranged so as to divide the available space equally. The other contained the same number of horizontal lines but these were arranged irregularly.

Experiment II investigated further the arrangement of the lines as a determinant of visual fixation in infants. There were six stimulus patterns (see Fig. 2). Three of these were composed of 5 lines and three were composed of 10 lines. Within each set of three, one of the patterns contained the specified number of horizontal lines, one contained a single vertical line bisecting the remaining horizontal lines perpendicularly, and the third pattern contained the number of vertical lines that maximized the number of blank spaces (three vertical and two horizontal for the five-line pattern and five vertical and five horizontal for the 10-line pattern). In each case the lines were arranged symmetrically, both in the horizontal and vertical dimensions.

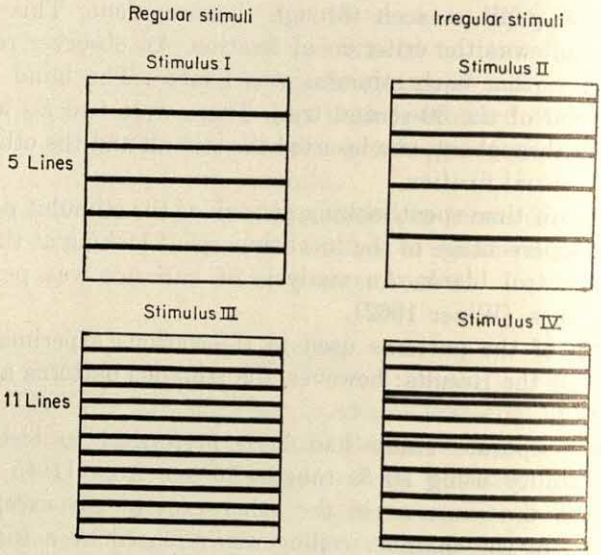


FIG. 1. Stimuli used in Exp. I.

Experiment III aimed to study whether the preference for patterns with crossings (Exp. II) is attributable merely to a preference for the vertical rather than horizontal lines. Three stimulus patterns were used in this experiment. The first was composed of five horizontal lines and was identical to stimulus 1 in Exp. II. The second was composed of

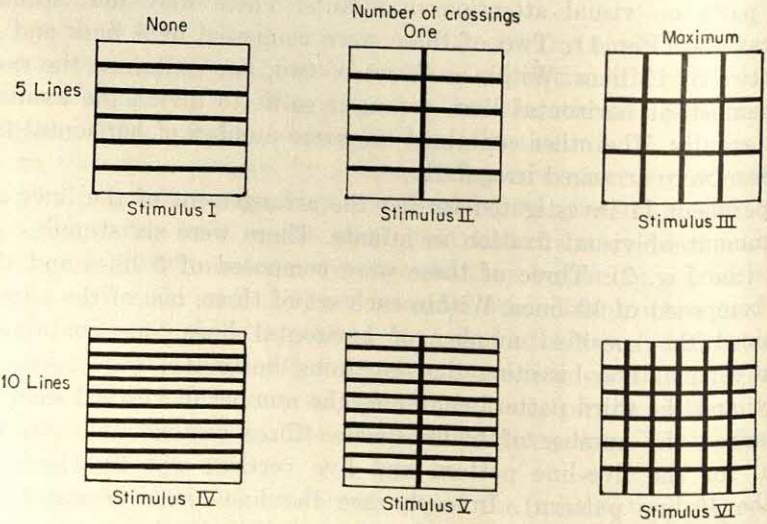


FIG. 2. Stimuli used in Exp. II.

five vertical lines and the third consisted of three vertical and two horizontal lines (stimulus 3, Exp. II).

The final study, Exp. IV, was designed to test whether the number of white rectangular spaces or "parts" was an important determinant of visual attention. There were six stimulus patterns. Three of these patterns consisted of 12 parts and three consisted of 16 parts. (See Fig. 3 for details.)

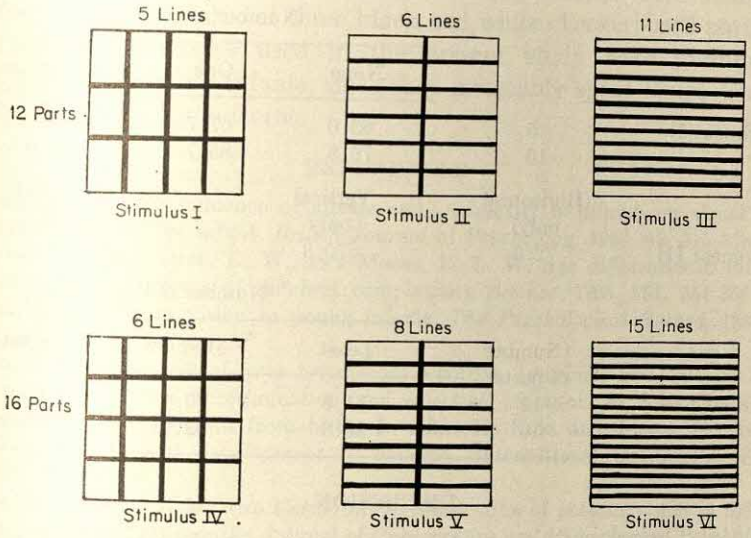


FIG. 3. Stimuli used in Exp. IV.

RESULTS

Analyses of variance were performed on the looking time percentages in all the experiments. Table 1 shows the mean percentage looking times for Exps. I–IV. In Exp. I, the infants were found to look significantly longer ($p < .05$) at stimulus patterns with 10 lines than at those with 5 lines. However, the other variable—irregularity of arrangement of the lines—was not significant. Experiment II showed that infants looked significantly longer at stimulus patterns with a greater number of lines ($p < .005$) and with a larger number of crossings because of the addition of the vertical line ($p < .025$). Experiment III showed that there was no significant difference in looking time between the patterns with five horizontal lines and that with five vertical lines but that both of these were looked at significantly less than was the pattern with three vertical and two horizontal lines ($p < .05$). The results of Exp. IV showed that the infants looked longer at those stimulus patterns with 16 parts (white rectangular spaces) than those with 12 parts ($p < .05$). The other variables—number of lines—was not significant in this instance.

TABLE 1
MEAN PERCENTAGE OF TOTAL LOOKING TIME FOR THE STIMULUS PATTERNS IN EXPS. I–IV

	Number of lines	Regular		Irregular
Experiment I	5	63.3		60.8
	10	82.4		80.8
		Number of crossings		
		None	One	Maximum
Experiment II	5	63.0	67.7	82.9
	10	76.8	85.7	86.0
	(Horizontal only)	Vertical only		Crossed vertical and horizontal
Experiment III	72.2	70.3		87.8
		Number of lines		
	(Number of parts)	Least	Medium	Most
Experiment IV	12	71.8	70.4	73.3
	16	86.3	82.2	85.3

DISCUSSION

In the present study each stimulus pattern was compared with a control blank, a feature which distinguishes this investigation from others in the field. This method of presentation allowed a more accurate assessment of the visual preferences of infants. In addition, the stimulus patterns were varied systematically in two ways: First, the number of lines, and second, the number of white rectangles or parts created by these lines. The infants looked longer at those stimuli with a greater number of lines, but when the number of parts was varied, the number of lines became a less important determinant of visual preference.

The results are in general agreement with some other workers in the field, notably Fantz (1966) who found that infants fixated patterned stimuli for longer than homogeneous ones. This is the case in the present experiment where the infant always looked longer at the patterned stimulus than at the simultaneously presented control blank. Also Berlyne (1958) and Fantz (1966) showed that certain patterns attracted more attention than others in infants and in fact these can be seen to contain a greater number of "parts."

However, it is important that the stimulus materials are not of so

fine a texture that the infant is unable to resolve them. Fantz, Ordy, and Udelf, (1962) have shown that the width of stripe that could be distinguished by infants decreased steadily with increasing age. Also Brennan, Ames, and Moore (1965) presented checkerboard patterns varying in the number of squares, and showed that the older the infants, the more complex was the checkerboard that they looked at longest. Such findings might well account for the discrepant results obtained by, for example, Hershenson (1964) where newborns were found to prefer the least complex of three black and white checkerboard patterns. The stimulus patterns used in the present study were presumably easily resolved by the infants, since they invariably spent longer looking at the most complex patterns.

REFERENCES

Berlyne, D. E. The influence of albedo and complexity of stimuli on visual fixation in the human infant. *British Journal of Psychology*, 1958, **49**, 315–318.

Brennan, W. M., Ames, E. W., and Moore, R. L. W. Age differences in infants' attention to patterns of different complexities. *Science*, 1965, **151**, 354–356.

Fantz, R. L. Pattern vision in young infants. *The Psychological Record*, 1958, **8**, 43–47.

Fantz, R. L. The origin of form perception. *Scientific American*, 1961, **204**, 66–72.

Fantz, R. L. Pattern discrimination and selective attention as determinants of perceptual development from birth. In Aline H. Kidd and Jeanne L. Rivoire (Eds.), *Perceptual development of children*. International Universities Press, New York, 1966.

Fantz, R. L., Ordy, J. M., and Udelf, M. S. Maturation of pattern vision in infants during the first six months. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 907–917.

Hershenson, M. J. Visual discrimination in the human newborn. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 270–276.

Saayman, G. S., Wardwell-Ames, E., and Moffett, A. M. Response to novelty as an indicator of visual discrimination in the human infant. *Journal of Experimental Child Psychology*, 1964, **1**, 189–198.

Spears, W. C. Assessment of visual preference and discrimination in the four-month-old infant. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 381–386.

Staples, R. The responses of infants to colour. *Journal of Experimental Psychology*, 1932, **15**, 119–142.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 180–189 (1969)

A Set to Establish Equivalence Relations in Pre-school Children[1]

ANN M. CLARKE, G. M. COOPER, AND ELIZABETH H. LOUDON
University of Hull, England

Sixty-four children, aged 3½-5 years, were given a single trial on a task demanding the association of categorically related nouns, presented in random order, to pictures representing human-beings, clothes, parts of the body, food, and tableware. They were allocated to one of three groups matched for age and low score, or to a group of higher age and score. Two of the three matched groups were trained for 10 trials on a task in which four categories (animals, furniture, vehicles, buildings) had to be differentiated. The same stimulus words were used in both training conditions, but in one case they were presented initially in categorically related clusters, while in the second condition presentation was invariably random. On transfer, *Ss* who had been trained with words presented in clusters showed a significant advantage over the other experimental *Ss* and both control groups. The results are interpreted as showing how conditions of training may modify a young child's natural style of response to a conceptual task and provide a short-cut for some part of his development of equivalence relations.

Previous experiments by the authors have shown (1) that normal pre-school children trained on complex tasks in which the categorization of geometric shapes was required to achieve success, showed a significantly higher level of performance on transfer to a task requiring categorization of pictures of common objects than children receiving either no training or training on a task where no categorization was required (Clarke & Cooper, 1966); (2) that the significant factor in inducing differential transfer was task complexity rather than task difficulty: Clarke & Cooper (1966) found no difference in transfer between two complex tasks which differed substantially in level of difficulty; Clarke, Cooper, & Clarke (1967) showed a very significant difference in transfer between two tasks both very difficult, but one involving no act of categorization; (3) that some degree of overlearning is another important factor in establishing an effective set. *Ss* trained to a specified criterion

[1] The writers are very grateful to the Association for the Aid of Crippled Children, New York, for the research grant which made this investigation possible. They are also indebted to the Chief Education Officers, Kingston-upon-Hull, and the East Riding of Yorkshire and their staffs for the provision of facilities in nursery schools.

of *one* high score only prior to transfer did less well than *Ss* who were trained over a larger number of trials (Clarke & Cooper, 1966); (4) that in the recall of a list of 16 words made up from four categories, randomly presented, adult mental retardates unexpectedly performed most efficiently if their training condition had involved memory for words presented in blocks containing categorically related material (Clarke, Clarke, & Cooper, 1968). These *Ss* achieved significantly higher scores on transfer than a group trained with unrelated words, and a small but probably reliable advantage over *Ss* trained with the same categorically related words randomly presented.

It was concluded that, in all these experiments, those *Ss* who showed an advantage on transfer had acquired a sensitivity to the categorical properties of the incoming stimulus material, which was not initially high in their behavioral repertoire.

It should follow from (3) and (4) above that if efficient performance on a transfer task demands the application of a conceptual rule, not readily available to the *Ss*, the most adequate training procedure will be one in which this rule is made rapidly apparent to him and is to some extent overlearned. Thus, instead of putting pre-school children through a lengthy training procedure (e.g. 24 trials) in which they had to discover for themselves the applicability of the principle of categorical equivalence, it might be more effective to present the material in such a way that the basis for correct responding was made clear.

The present experiment was designed to test the hypothesis that if a limited number of training trials (e.g. 10) are given at the rate of two per day, *Ss* who have material presented in categories should have an advantage over *Ss* trained with the same material randomly presented, both at the end of training and on transfer to a new task. It was further predicted that both groups should show an advantage over a matched group which had received no training.[2]

METHOD

The Tasks

The *transfer task* consisted of 25 words, five in each of five conceptual categories, presented randomly, to be assigned by the *S* to one of five

[2] We are grateful to the reviewer of this paper who pointed out the relevance to this study of the experimental procedure and results of Kurtz and Hovland (1956), and the discussion by Bourne (1966, p. 61) of these and other findings. These investigators found that concept learning was accelerated when exemplars of a concept were presented in close temporal proximity rather than separated from one another (i.e., unmixed presentation resulted in faster learning than any mixed order).

pictures, each representing a different category: people, parts of the human body, clothes, food, and tableware. The pictures (black outline drawings on 3 × 3-inch white cards) were: a fully clothed man, a pair of trousers, an eye, a cake, and a cup and saucer. The words were: head, egg, girl, fork, meat, cap, jug, dress, doctor, shoes, apple, finger, jelly, baby, tongue, bread, plate, man, knee, gloves, spoon, dish, teacher, vest, foot.

The *training task* consisted of words from four other categories (vehicles, furniture, animals and buildings) with a picture representing each: a boat, a stool, a rat, and a house. Lists of words (to be assigned to the pictures) were drawn up; these varied slightly in length, and also in content from trial to trial, although the categories remained constant.

Two training conditions were used: (1) *Blocks* (B). The words were presented in conceptually related clusters over two trials, followed by a gradual reduction in cluster size, until on the last two of ten training trials no two related words occurred together. Thus, the list for trial 1 was: pig, mouse, horse, bed, stool, desk, car, boat, aeroplane, dog, bear, hut, church, garage, school, bicycle, train, wagon, rocket, chair, table, cupboard (22 words); for trial 5 was: settee, cupboard, stool, bed, rabbit, table, rocket, lorry, aeroplane, horse, dog, train, hut, car, wagon, tiger, pig, mouse, desk, wardrobe, school, church, bicycle, (23 words); and for trial 10 was: station, dog, bicycle, shop, stool, bear, desk, tiger, house, wagon, wolf, shed, rabbit, school, lorry, church, wardrobe, garage, aeroplane, settee, rocket, hut, train, (23 words). (2) *Random Category* (RC). The same words as for the B group were used on each trial of training, except that from the start no two conceptually related words occurred in juxtaposition.

Procedure

The apparatus consisted of a presentation panel on which the stimulus pictures were positioned. Beneath each picture was a button which had to be pressed by the *S* to indicate his choice of stimulus when presented with a word given orally by *E*. The positions of the pictures on the panel were changed each day during training and transfer.

The two experimental groups were given one trial pretest on the transfer task. After 2 days had elapsed, they were trained on their respective conditions for ten trials, (two trials per day) then after a further 2 days eight trials on the transfer task were given (two trials per day).

Two control groups were treated in the same way except that they received no training condition between pretest and eight trials of trans-

fer. Care was taken to maintain good rapport with these children throughout the experiment.

Before pretest all the *Ss* were given a "warm-up" period on the apparatus so that an efficient response was established. Color names were given orally by *E* one at a time, each *S* then matched to a plain color picture (red, blue, green, yellow) by pressing the appropriate button. Two trials on color/word match were given. Two further trials on object/word match were given. The pictures were fish, drum, tree, bell; the words were the names of the represented objects.

The instructions for these tasks were:

> "I am going to say some words. When I say a word, I want you to show me which picture the word goes with by pressing the button beneath the correct picture. Are you ready?"

All the *Ss* readily performed these warm-up tasks. Instructions on the pretest and throughout training and transfer were the same as above. When an error was made, *E* said "No" and repeated the word. In the event of three errors with one word, he demonstrated the appropriate response. A record was kept of the nature of all incorrect responses.

Subjects

The *Ss* were children attending public nurseries, places at which are granted on the basis of social need *either* (1) for reasons of domestic overcrowding, lack of play facilities in the home, or inability of the mother to give the child adequate attention; *or* (2) because the mother is in regular employment, often in a profession such as nursing or teaching for which there is a shortage of personnel. Children who were mentally retarded or suffering from abnormalities of speech or hearing were excluded.

A group of forty-three children aged 3½ years to just under 5 years were given the warm-up procedure already described followed by a single trial on the transfer task (T). Scores on T ranged from 5 to 20 correct out of a maximum possible of 25.

Thirty children were selected from the total group according to the following criteria: (a) That their score (i.e., the number of first correct responses) did not exceed 10/25 words correct; (b) that their score on the latter part (last 12 words) of T was not higher than that for the first 13 words—in effect those who seemed least likely to have the set readily available. The thirty children were then allocated to one of three groups so that there should be a close match for age, sex, and score on T. The resulting three groups of ten children were allocated randomly to B training, RC training, or no training (Matched Control

group, MC). Of the remainder, the three with the highest scores were excluded altogether and the remaining ten children formed the High Control group (HC) receiving no training, but being given the transfer task.

The results were consistent with the prediction. On trial 1 of transfer the B group was significantly superior to the other three groups, including HC; thereafter, they remained above the level of the RC and MC groups, while the HC *Ss* reached a level equivalent to the B group.

It seemed wise to repeat the study "blind." When further *Ss* became available in another kindergarten, thirty-six children were pretested and assigned on the basis of score, age, and sex to one of six groups, with six *Ss* in each: B, RC, MC, and HC (as before), High B and High RC (new groups). The training was carried out by one author (A.M.C.), and the transfer testing by another (G.M.C.) who had no knowledge of any *S*'s group assignment. The results showed clearly a precise repetition of the previous trends; thus, the data for the two studies were pooled, giving four groups each with sixteen *Ss*. The results for the two new groups of six *Ss*—High B and High RC—were consistent with expectation, but will not be presented since the major reason for including these children was to control possible *E* effects during transfer.

RESULTS

Pretest Data

Three groups, each of sixteen *Ss*—B, RC, and MC—were matched on pretest; i.e., on one trial of the transfer task and on age ($F = .14$, $df = 2, 45$, N.S.). The fourth group, HC, had a pretest score very significantly higher than the other three groups and its average age was slightly higher. The mean age per group in months was: B, 50.81; RC, 49.75; MC, 50.50; HC, 54.19. There was a significant correlation between pretest score and age ($r = .373$, $df = 63$, $p < .01$).

The mean scores for each group at pretest together with their scores per category, out of a maximum possible of 5, are shown in Table 1. A two-way analysis of variance indicated that the three matched groups were well balanced with respect to category scores, and that there were large differences in relative difficulty among the categories (Pretest: F for three matched groups $= .002$, $df = 2, 225$; F for categories $= 57.00$, $df = 4, 225$, $p < .001$; Interaction $F = 1.19$, $df = 8, 225$, N.S.). Addition of the HC group resulted in a significant groups difference ($F = 11.84$, $df = 3, 300$, $p < .001$), a significant categories term ($F = 71.92$, $df = 4, 300$, $p < .001$) and an interaction of <1. It was concluded, therefore, that the groups were consistent with respect to category difficulty, and

that the older group was more efficient than were the three younger groups.

Data on Training (B and RC groups)

The average scores for each of the two groups per trial of training are recorded in Figure 1. It should be remembered that the number of items presented during training varied from trial to trial, although on all trials the items were the same for both groups. Members of the two groups performed similarly on the first two trials of training ($F = <1$ for trials 1 and 2), and it was assumed that at this stage the B *Ss* failed to perceive, or to utilize, the categorical relations between the words presented to them in blocks. Thereafter, this group developed an advantage, and by the last trial of training, when the words were randomly ordered for both groups, the difference in score was significant ($F = 6.54$, $df = 1, 30$, $p < .025$). In this connection, see Footnote 2.

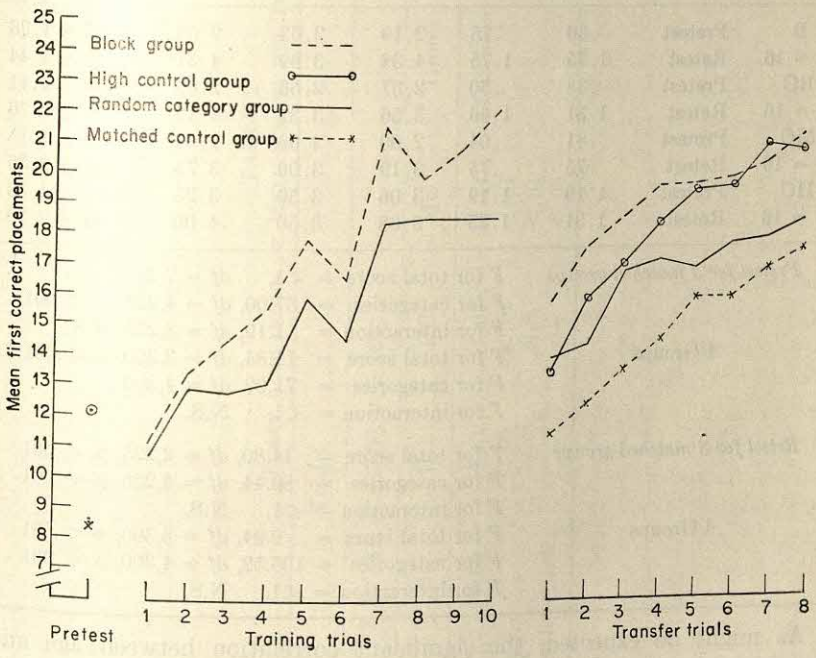


FIG. 1. Graphs showing: (1) Average correct responses at pretest for four groups on the transfer task; (2) average scores per trial on training for the two experimental groups—B and RC. The number of stimulus words per trial were: 22, 21, 21, 21, 23, 20, 24, 24, 23, 23, to be associated into four categories; (3) average scores per trial (out of 25) on the transfer task comprising five words for each of five categories.

Data on Transfer

Table 1 shows the mean score for each group at retest, together with their scores per category. A two-way analysis of variance for the three matched groups gave: F for groups = 14.80, $df = 2, 225$, $p < .001$; F for categories = 80.44, $df = 4, 225$, $p < .001$; interaction F is <1). A similar analysis for the two experimental groups showed a significant superiority of the B over the RC group (F for groups = 5.94, $df = 1, 150$, $p < .025$; F for categories = 53.16, $df = 4, 150$, $p < .001$). The B group on retest was significantly superior to the HC group, who had been selected for their superiority to other *Ss* on pretest.

TABLE 1

MEAN SCORES (number of correct first responses out of a maximum of 5) PER CATEGORY FOR THE FOUR GROUPS AT PRETEST AND RETEST (trial 1 of transfer). MEANS AND STANDARD DEVIATIONS OF TOTAL SCORES (out of 25) ARE ALSO GIVEN

Group	Category	Clothes	Body	Food	Tableware	People	Mean total
B	Pretest	.50	.75	2.19	2.63	2.50	8.57 ± 1.26
$N = 16$	Retest	1.75	1.75	4.38	3.69	4.31	15.88 ± 2.44
RC	Pretest	.38	.50	2.37	2.56	2.69	8.50 ± 1.41
$N = 16$	Retest	1.31	1.56	3.56	3.32	4.19	13.94 ± 2.76
MC	Pretest	.81	.94	2.00	2.06	2.63	8.44 ± 1.15
$N = 16$	Retest	.75	.75	3.19	3.00	3.75	11.44 ± 2.30
HC	Pretest	1.19	1.19	3.06	3.50	3.25	12.19 ± 1.33
$N = 16$	Retest	1.31	1.25	3.38	3.50	4.00	13.44 ± 2.70

Pretest for 3 matched groups	F for total score = <1, $df = 2,225$
	F for categories = 57.00, $df = 4,225$, $p < .001$
	F for interaction = 1.19, $df = 8,225$, N.S.
4 Groups	F for total score = 11.84, $df = 3,300$, $p < .001$
	F for categories = 71.92, $df = 4,300$
	F for interaction = <1, N.S.
Retest for 3 matched groups	F for total score = 14.80, $df = 2,225$, $p < .001$
	F for categories = 80.44, $df = 4,225$, $p < .001$
	F for interaction = <1, N.S.
4 Groups	F for total score = 9.94, $df = 3,300$, $p < .001$
	F for categories = 105.52, $df = 4,300$, $p < .001$
	F for interaction = <1, N.S.

As might be expected, the significant correlation between age and score, present at pretest, was not replicated at retest ($r = .113$, N.S.) nor was there a correlation between age and summed score over the first four trials of transfer ($r = .172$, N.S.).

The average score for each group per trial on transfer is shown in Figure 1. *Ss* in the B group maintained their superiority over the two

TABLE 2
TUKEY COMPARISON OF MEAN SCORES FOR TRIAL 4 AND TRIAL 8

	Trial 4				Trial 8				
	$F = 9.61$, $df = 3,60$, $p < .001$				$F = 5.78$, $df = 3,60$, $p < .001$				
Group		B	HC	RC	MC	Group	B	HC	RC	MC
Mean score		19.69	18.31	17.31	14.44	Mean score	21.25	20.81	18.25	17.44
	B	—	$< .05$	$< .01$	$< .01$	B	—	N.S.	$< .05$	$< .01$
	HC		—	N.S.	$< .01$	HC		—	$< .05$	$< .01$
	RC			—	$< .05$	RC			—	N.S.

matched groups during the eight transfer trials, but were overtaken by the HC group. Table 2 shows the results of a comparison of the mean scores for the four groups at Trial 4 and Trial 8, using Tukey's method of multiple comparisons (Ryan, 1959).

Thus the data clearly indicate a consistent superiority of the B group over the RC and MC groups during the eight trials of transfer, and an initial superiority over the HC *Ss*, whose rate of learning, however, was sufficiently accelerated to achieve scores commensurate with the B group in the latter part of the experiment.

An examination of the errors made by the three matched groups on transfer showed that the differences between them were quantitative rather than qualitative and that the areas of confusion were similar. There was a strong tendency on the part of control *Ss* to associate clothes and parts of the body with the human figure, and to confuse tableware and food. Although they made fewer errors, members of the experimental groups showed a similar pattern of erroneous responses.

DISCUSSION

Performance on the transfer task, which is in any degree efficient, implies an ability to discriminate between closely related categories, and to establish equivalence for different exemplars within categories on the basis of superordinate structures, which are sufficiently available to the *S* for action though not necessarily for verbal formulation.

It seems reasonable to assume from the pretest scores that this style of responding is not high in the behavioral repertoire of the normal child in his fourth and fifth years of life, although the occurrence of a few very high scores, and the significant correlation between pretest score and age provides evidence that it is in the process of developing. The consistency of the groups at pretest with respect to difficulty of categories, and the similarity in type of erroneous response on transfer provides evidence for a cognitive strategy, albeit inadequate, on the part of control *Ss* and against their use of a policy of random responding.

The effect of training on a transfer task may be measured in two ways: (1) Quality of performance on a single trial of a new task and (2) efficiency of learning on repeated trials of that task; conclusions concerning what is transferred will depend upon the extent to which *Ss* show immediate gains or accelerated learning, or both. In this case the B group was significantly superior to the other three groups on immediate transfer, and thereafter consolidated its position, attaining high scores at the end of training. The HC group, starting from a position significantly inferior to B *Ss*, accelerated and after four trials the difference between them was reduced to non-significance. The second experimental

group (RC) had a significant advantage over the MCs and were equal to the HC group on trial 1, but unlike them, failed to maintain their advantage, so that the gap between RC and MC *Ss* diminished until there was no significant difference in score.

The evidence suggests a general acquaintance with the categories prior to the experiment, common to all *Ss*, and that, as predicted, the B group acquired early during training the idea of equivalence relations, enjoying sufficient experience with it in one situation to apply it immediately to another task. By contrast, members of the RC group (the whole training procedure of which was identical with that of B, excepting only the categorical ordering of the words for the latter group) had at the end of the relatively brief training period made only some progress towards discovering the principle, but had not sufficiently learned or overlearned the rule for it to be reliably used in the transfer situation.

The results of this experiment may be interpreted as indicating that it is possible to accelerate the normal "maturational" process of rule-forming behavior by imposing certain conditions of training, and that it is important to evaluate the relative efficacy of several programs. The findings are consistent with the view of theorists who, when interpreting findings on the maturation of intelligent behavior, insist on the importance of the antecedent conditions of learning and reinforcement.

REFERENCES

Bourne, L. E., Jr. *Human conceptual behavior.* Pp. 60–62. Boston: Allyn and Bacon, 1966.

Clarke, A. M., Clarke, A. D. B., and Cooper, G. M. The development of a set to perceive categorical relations. *In* H. C. Haywood (Ed.), *Social-cultural aspects of mental retardation.* New York: Appleton (in press).

Clarke, A. M., and Cooper, G. M. Transfer in category learning of young children: its relation to task complexity and overlearning. *British Journal of Psychology,* 1966, **57**, 361–373.

Clarke, A. M., Cooper, G. M., and Clarke, A. D. B. Task complexity and transfer in the development of cognitive structures. *Journal of Experimental Child Psychology,* 1967, **5**, 562–576.

Kurtz, K. H., and Hovland, C. I. Concept Learning with differing sequences of instances. *Journal of Experimental Psychology,* 1956, **51**, 239–243.

Ryan, T. A. Multiple comparisons in psychological research. *Psychological Bulletin,* 1959, **56**, 26–47.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 190–194 (1969)

R-S Learning in Children[1]

DONALD H. KAUSLER AND MARY ANN GOTWAY[2]
Saint Louis University

R-S learning for kindergarten, second, fourth, and sixth grade *Ss* was measured following the attainment of one perfect S-R trial on a paired-associate list (six pairs, pictures as stimulus and response items). A significant increment in S-R learning proficiency was found between grade groups. However, differences between grades in mean R-S recall scores were negligible, and the amount of R-S learning was approximately 80% of S-R learning for each grade level. These results were viewed as evidence against an incidental learning interpretation of R-S learning.

The phenomenon of R-S, or "backward," learning during S-R practice on a paired-associate task has considerable importance for research and theory on developmental trends in verbal learning processes (Goulet, 1968). Of special importance is the possible effect of differential amounts of R-S learning that may occur with increasing age on overall transfer effects for paradigms presumed to involve interference effects via competing R-S associations (e.g., the A-B, C-B paradigm; Martin, 1965).

Nevertheless, R-S learning in children has received little empirical attention. In what is apparently the only study measuring R-S learning in children, Palermo (1961) reported no difference in amounts for fourth and sixth grade *Ss*. In addition, the relative amount of R-S learning, that is, relative to the amount of S-R learning, was considerably less than the relative amount typically demonstrated for college age *Ss*—57% (Palermo, 1961) vs. 80% or more (e.g., Dron and Boe, 1964)—when well-integrated items serve as stimulus terms. However, the transfer methodology employed in the Palermo study may well have masked both an increment in R-S learning from the fourth to sixth grade levels and a larger amount of R-S learning relative to S-R learning at both grade levels. The methodology involved a mixed transfer list in which there were two A-B pairs (i.e., S-R pairs from the first list), two B-A pairs (i.e., R-S pairs from the first list), and two C-D pairs (i.e., new

[1] The authors are indebted to the principal and teachers of St. Rochs School, St. Louis, Mo. for the procurement of subjects in this research.

[2] This research was conducted during the tenure of the second author as a National Defense Education Act Fellow in psychology.

stimulus and response terms). R-S learning was measured in terms of correct anticipations of B-A pairs on the transfer list relative to correct anticipations of A-B pairs. Mixed transfer lists are susceptible to rehearsal strategies (Paul and Paul, 1968) which conceivably favor S-R pairs over R-S pairs. Moreover, the operation of such strategies may vary with developmental level, thereby masking differential amounts of R-S learning *per se*. The present study investigated R-S learning by means of the standard R-S recall trial following attainment of mastery of the list in the S-R direction. In addition, the age range tested was extended by including kindergarten and second grade *S*s, as well as fourth and sixth grade *S*s.

METHOD

Subjects. The *S*s consisted of 64 children, 32 boys and 32 girls, randomly selected from the kindergarten, second, fourth, and sixth grades ($n = 16$, 8 boys and 8 girls at each grade level). The school was a parochial school located in a middle class St. Louis neighborhood.

Lists. The stimulus and response terms were simple line drawings selected from below the 3-year 6-month age level on the Peabody Picture Vocabulary Test. An initial pool of 24 pictures was selected such that the objects represented seemed to be semantically and associatively unrelated to one another. These pictures were then divided by random assignment into two sets of 12 pictures each. A separate list of six pairs was formed from each set by the random selection and pairing of six pictures as stimuli and six pictures as responses. Two separate lists were employed for purposes of greater generalization of findings and control over idiosyncratic associations. Half of the *S*s at each grade level received one list, and the other half received the second list. Both S-R learning and R-S recall differentials between lists were negligible across all grade levels, and the two lists were collapsed for the analyses reported below.

Procedure. Prior to the experiment proper, each *S* was given several trials on a two-pair practice list to insure comprehension of the instructions. In addition, prior to the start of practice on the experimental list each *S* was shown, on individual sheets of paper and in a randomized order, each of the 12 pictures used in the list and was asked to name the object depicted. This procedure made certain that the stimulus and response terms were within *S*'s active vocabulary.

The experimental list was presented for S-R practice by the anticipation method, with a 4:4-second rate of exposure and an 8-second intertrial interval. As a control for serial cues, three different random orders of pairings were presented across trials. Practice continued to a criterion

of one perfect trial. The R-S recall measure followed immediately after the attainment of the S-R criterion. For R-S recall each R item was exposed alone for 4.5-seconds, and *S* was asked to name the picture that had been paired with it. Two different random orders of R item presentation were employed, with half of the *S*s in each grade receiving one order, etc. Order of presentation was found to have a negligible effect on the amount of R-S recall across grade levels, and the two orders were collapsed for the analyses reported below. For both S-R and R-S trials items were projected on a Telescreen by means of a Carousel slide projector.

RESULTS AND DISCUSSION

For trials to S-R criterion the means and standard deviations were 10.00 and 4.49, 8.68 and 3.28, 7.06 and 4.26, and 7.12 and 5.22 for kindergarten, second, fourth, and sixth grade groups, respectively. The medians for trials to criterion were 10.00, 7.50, 5.50, and 5.00 in order of increasing grade level. The disparity between the mean and the median for both the fourth and sixth grade groups, together with the large coefficient of variation apparent in both groups, reflects the pronounced positive skew present in the distribution of trials scores for both groups. Consequently, differences between central tendencies were tested for statistical significance by means of the median test. The overall between-group effect was significant, $X^2(3) = 11.68$, $p < .01$. In addition, the contrast between the combined kindergarten and second grade groups and the combined fourth and sixth grade groups was significant, $X^2(1) = 7.68$, $p < .01$. The contrast between the kindergarten and second grade groups approached significance, $X^2(1) = 3.33$, $p < .10$, but the contrast between fourth and sixth grade groups was obviously not significant, $X^2(1) < 1$. These results support the usual finding (e.g., Jensen and Rohwer, 1965) of increasing S-R learning proficiency from kindergarten through the fourth grade, but with little change in proficiency from the fourth to the sixth grade.

Means and standard deviations for R-S recall were 5.06 and 1.18, 4.75 and 1.00, 5.25 and .86, and 4.81 and 1.17 for kindergarten, second, fourth, and sixth grade groups, respectively. An analysis of variance substantiated the apparent absence of a difference in R-S recall scores between grade levels, $F(3, 60) < 1$. Although acceptance of the null hypothesis is, of course, impossible, the present results nevertheless strongly imply that differences in R-S learning from the kindergarten through the sixth grade level are, at most, slight. A complicating factor in accepting this view is

the evidence that the groups did differ in the number of trials required to attain the S-R criterion. Conceivably, increases in the number of S-R trials may provide an additional opportunity to learn R-S associations, thereby serving to inflate R-S scores in the younger age levels. This hypothesis suggests the presence of a positive correlation between S-R trials and amount of R-S recall. However, the within-group rank-order correlation coefficients for the present groups were all negligible and statistically nonsignificant (rho = .05, .18, —.34, and —.15 in order of increasing grade level; all p values $>.10$), and the differential number of trials between groups seems to be an unlikely confounding factor.

The failure to find a difference in R-S learning between fourth and sixth grade Ss is in agreement with Palermo's (1961) earlier finding. However, the present results depart from those of Palermo in estimating the amount of R-S learning in children relative to S-R learning. As indicated earlier, this estimate was 57% in the Palermo study, whereas the present values (R-S score/6 $\times$ 100) ranged from 79 to 87%. These higher estimates are in close agreement with the values found in studies employing young adults (e.g., Dron and Boe, 1964).

The absence of increments in R-S learning with increasing age may be taken as evidence against an incidental learning interpretation of R-S phenomena (cf. Ekstrand, 1966, for a discussion of the incidental learning conception of R-S phenomena). Goulet (1968) has noted that studies with other incidental learning tasks (e.g., Siegel and Stevenson, 1966) have yielded monotonic increments in amounts of incidental learning through the sixth grade level, and that a similar function is to be expected for R-S learning if it is truly a variant of incidental learning. However, further research is obviously needed before definitive answers can be reached regarding the nature of R-S learning processes in children and the nature of potential changes in these processes with increasing age. Especially valuable research would be that which would involve bidirectional practice sequences (Goulet, 1968).

REFERENCES

Dron, D. M., and Boe, E. E. R-S paired associate learning as a function of percentage occurrence of response members. *J. of Verbal Learning and Verbal Behavior,* 1964, **3,** 130–131.

Ekstrand, B. R. Backward associations. *Psychological Bulletin,* 1966, **65,** 50–64.

Goulet, L. R. Verbal learning in children: Implications for developmental research. *Psychological Bulletin,* 1968, **69,** 359–376.

Jensen, A. R., and Rohwer, W. D., Jr. Syntactical mediation of serial and paired-associate learning as a function of age. *Child Development,* 1965, **36,** 601–608.

Martin, E. Transfer of verbal paired associates. *Psychological Review,* 1965, **72,** 327–343.

PALERMO, D. S. Backward associations in the paired-associate learning of fourth and sixth grade children. *Psychological Report,* 1961, **9,** 227–233.

PAUL, C., AND PAUL, H. Transfer-activated response sets in verbal learning and transfer. *Psychological Review,* 1968, **75,** 537–549.

SIEGEL, A. W., AND STEVENSON, H. W. Incidental learning: A developmental study. *Child Development,* 1966, **37,** 811–818.

Reward and Punishment in Human Learning

Elements of a Behavior Theory

By **Joseph Nuttin**
Department of Psychology
University of Louvain
Louvain, Belgium

In collaboration with
Anthony G. Greenwald
Department of Psychology
Ohio State University
Columbus, Ohio

A psychological interpretation of the operation of reward and punishment in the shaping of memory and behavior. The conclusions arrived at through the authors' numerous experiments are strikingly different from those obtained by Thorndike and others working in this field. The research focuses particularly on the role of the learner's task orientation when receiving reward or punishment, the effects of varying the ratio of reward to punishment in the structure of a task, and the determinants of memory for reward and punishment. Discussions of methodological problems associated with research on reward and punishment are also presented.

1968, 205 pp., $8.50

AP 2273

Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

Negro and White Children

A PSYCHOLOGICAL STUDY IN THE RURAL SOUTH

By **E. Earl Baughman**
W. Grant Dahlstrom
Department of Psychology, University of North Carolina Chapel Hill, North Carolina

1968, 572 pp., $15.00

Reports the scientific conclusions of a unique four year study of racially-segregated students in a rural, economically deprived area of North Carolina.

CONTENTS:

SETTING AND METHODS
Life in Millfield
Procedures Used to Study Millfield's Children
INTELLIGENCE AND ACADEMIC ACHIEVEMENT
Ability Differences between Racial, Sex, and Age Groups
Academic Achievement Differences between Racial, Sex, Age, and Ability Groups
The Relationship of Selected Family Variables to Ability and Academic Achievement.
INTELLECTIVE CHANGE
Changes in IQ Among Young Negro Children
Changes in IQ Among White Adolescents
The Effect of Kindergarten Training upon the Intellective Functioning of Negro and White Children
A Pilot Project To Accelerate the Intellective Development of Four-Year-Old Negro and White Children
PERSONAL ATTRIBUTES
Millfield's Teachers View Their Pupils
Eighth-Grade Children as Assessed by the Minnesota Multiphasic Personality Inventory
Fantasy Behavior: Achievement Motivation
Fantasy Behavior: Other Attributes
Correlates of Peer Popularity
PERSONAL PERSPECTIVES
Mothers of Kindergarten-Age Children Report on Their Children, Training Practices, and Families
Eighth-Grade Children Report on Their Worlds
INTEGRATION AND RECOMMENDATIONS
Some Summary Comments and Observations
Narrowing the Gap

REFERENCES.
APPENDIX
AUTHOR INDEX—SUBJECT INDEX.

AP 2155

Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

INFORMATION FOR AUTHORS

The *Journal of Experimental Child Psychology* will publish papers in which the behavior and development of children is clearly related to its determining variables. Typically (not always) this would mean that a variable has been manipulated in an experimental manner. Articles will be reviewed in terms of the apparent reliability of their findings and the clarity of the relationships they affirm. Research conceived in any theoretical framework is invited, so long as its conclusions are reliable and compelling. Occasionally, critical reviews of related bodies of experimental literature, theoretical discussions which define new paths of experimental work, and well-controlled developmental studies on infra-human subjects will be published. The criterion for acceptance of such papers will be their value in prompting experimental research in related areas with children. Short notes describing new apparatus or techniques, new materials, circuits, or other information valuable to investigators in the field will be accepted from time to time. There is no page charge for publication.

Two copies of each manuscript and each figure or table should be sent to:

Dr. Sidney W. Bijou, Child Behavior Laboratory, 403 E. Healey, University of Illinois, Champaign, Illinois 61820

A third copy should be retained by the author. Manuscripts must be double-spaced and may be typed, dittoed, or mimeographed. Each manuscript must include a concise abstract, preferably not exceeding 300 words, which will appear after the title and author's name in the published article. In some papers this abstract will obviate the need for a summary. In general, authors are free to use whatever organization their paper seems to require, but the specifications of the *Publication Manual* of the American Psychological Association are strongly recommended.

Figures and Tables. Figures should be numbered consecutively with Arabic numerals in order of mention in the text; each figure should have a descriptive legend. Legends should be typed together on a separate sheet, double-spaced. All illustrations should be in finished form ready for the engraver. A convenient size for drawings is 8½ × 11 inches (21 × 27.5 cm). Drawings should be made with India ink on tracing linen, smooth surface white paper, or bristol board. Graph paper if used should be ruled in blue. Grid lines that are to show in the final engraving should be inked in black. Sharp, clear glossy prints of figures are acceptable.

Tables should be typed on separate pages, numbered consecutively with Arabic numerals in order of mention in the text. All tables should have descriptive headings, typed (double-spaced) above the table.

Literature references in the text should be cited by author's surname and the year of publication, e.g., Smith, 1960; Smith *et al.*, 1959; Smith and Jones, 1961, p. 250 (for references to a specific page). If a reference has more than two authors, the citation includes the surnames of all authors at the first mention, but later citations of the same reference include only the surname of the first author and the abbreviations *"et al."* Suffixes *a, b,* etc., should be used following the date to distinguish two or more works by the same author(s) in the same year, e.g., Smith, 1960a, 1960b. Literature citations in the bibliography should be arranged alphabetically according to the surname of the author.

Journal titles should be written out in full according to the form followed in the Psychological Abstracts (Vol. 39).

The desired style of capitalization and punctuation in references is apparent in the following examples:

Staats, A. W., and Staats, Carolyn K. A comparison of the development of speech and reading behavior with implications for research. *Child Development,* 1962, 33, 831–846.

Spiker, C. C. Research methods in children's learning. In P. H. Mussen (Ed.), *Handbook for research methods in child development.* New York: Wiley, 1960. Pp. 374–420.

Munn, N. L. *The evolution and growth of human behavior.* Cambridge: Riverside Press, 1955.

Reprints: Fifty gratis reprints are supplied and additional copies may be ordered.

Computer-Assisted Instruction:

Stanford's 1965-66 Arithmetic Program

By **Patrick Suppes**
Max Jerman
Dow Brian
Institute for Mathematical Studies in the Social Sciences
Stanford University, Stanford, California

In collaboration with **Diana Axelsen, Guy Groen, Lester Hyman,** and **Brian Tolliver**

CONTENTS:

Introduction
Description of the Project
Edited Daily Log
Student, Parent, and Teacher Questionnaires
Survey of Previous Research on Arithmetic
Performance Models for Addition, Subtraction, and Multiplication
Performance Models for Fractions, Division, and Mixed Drills
An Experiment on the Commutative, Associative, and Distributive Laws
The Time-Sharing System and CAI Programming
Appendix I: Computer-Based Mathematics Instruction—Second Year of the Stanford Project (September 1, 1964—August 31, 1965)
Appendix II: Some Sample Drills
REFERENCES—SUBJECT INDEX.

1968, 385 pp., $7.50

AP 2175

Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

JOURNAL OF Experimental Child Psychology

EDITOR: Sidney W. Bijou

ASSOCIATE EDITORS: Richard S. Bogartz
David Zeaman

EDITORIAL BOARD:

Donald M. Baer
Albert Bandura
Wesley C. Becker
Harry Beilin
Lyle E. Bourne, Jr.
Gordon N. Cantor
Charles C. Clifton, Jr.
Peter Eimas
Robert L. Fantz
Eleanor J. Gibson
Eugene S. Gollin
Willard W. Hartup
Raymond H. Hohle
Herbert Kaye
Tracy S. Kendler
Helen L. Koch
Eugene R. Long
Harry Munsinger
David S. Palermo
Ross Parke
Herbert L. Pick
Hayne W. Reese
Todd Risley
Bryan E. Shepp
Murray Sidman
Einar Siqueland
Patrick Suppes
Paul Weisberg
Joachim Wohlwill
Betty House Zeaman

Volume 8, Number 2, October 1969

ACADEMIC PRESS
New York and London

CONTENTS

TOM TRABASSO, MARY STAVE, AND ROBERT EICHBERG. Attitude Preference and Discrimination Shifts in Young Children 195

LUDWIG IMMERGLUCK AND MARIA CLAUDIA MEARINI. Age and Sex Differences in Responses to Embedded Figures and Reversible Figures . . . 210

PAUL WEISBERG AND DALE B. KENNEDY. Maintenance of Children's Behavior by Accidental Schedules of Reinforcement 222

FRANK H. HOOPER. Piaget's Conservation Tasks: The Logical and Developmental Priority of Identity Conservation 234

BRENDA BRYANT AND MOSHE AINSFELD. Feedback verses No-Feedback in Testing Children's Knowledge of English Pluralization Rules . . . 250

KATHERINE J. NELSON, TAMAR ZELNIKER, AND WENDELL E. JEFFREY. The Child's Concept of Proportionality: A Re-examination 256

TORU NAITO AND LEWIS P. LIPSITT. Two Attempts to Condition Eyelid Responses in Human Infants 263

HARRIET L. RHEINGOLD AND CAROL O. ECKERMAN. The Infant's Free Entry into a New Environment 271

KATHERINE J. NELSON. The Organization of Free Recall by Young Children 284

CORNELIA DODD AND MICHAEL LEWIS. The Magnitude of the Orienting Response in Children as a Function of Changes in Color and Contour . 296

MICHAEL D. ZEILER AND CRYSTAL A. KELLEY. Fixed-Ratio and Fixed-Interval Schedules of Cartoon Presentation 306

GRAEME S. HALFORD. An Experimental Analysis of the Criteria Used by Children to Judge Quantities 314

JUAN PASCUAL-LEONE AND JUNE SMITH. The Encoding and Decoding of Symbols by Children: A New Experimental Paradigm and a Neo-Piagetian Model 328

DAVID C. SPEER, PETER F. BRIGGS, AND ROCHELLE GAVALAS. Concurrent Schedules of Social Reinforcement and Dependency Behavior among Four-Year-Old Children 356

THOMAS J. TIGHE AND LOUISE S. TIGHE. Facilitation of Transposition and Reversal Learning in Children by Prior Perceptual Training . . . 366

LEONARD M. HOROWITZ, ANITA K. LAMPEL, AND RUBY N. TAKANISHI. The Child's Memory for Unitized Scenes 375

FRANK D. TURRISI AND BRYAN E. SHEPP. Some Effects of Novelty and Overtraining on the Reversal Learning of Retardates 389

REGINA A. MODRESKI AND ALBERT E. GROSS. Young Children's Initial and Changes Names for Form-Color Stimuli 402

ERRATA 410

Published bimonthly at Exchange Place, Hanover, Pa. 17331 by Academic Press, Inc., 111 Fifth Avenue, New York, N. Y., 10003.

In 1969, Volumes 7 and 8 (consisting of 3 issues each) will be published.

Price: $22.00

Private subscriptions (for subscriber's personal use): $10.00

In 1970, Volumes 9 and 10 (consisting of 3 issues each) will be published.

Price $24.00

Private subscriptions (for the subscriber's personal use): $10.00

All correspondence and subscription orders should be addressed to the office of the Publishers at 111 Fifth Ave., New York, N.Y. 10003

Send notices of change of address to the office of the Publishers at least 4 weeks in advance.

Please include both old and new addresses.

Second-class postage paid at Hanover, Pennsylvania 17331

© 1969 by Academic Press, Inc.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 195–209 (1969)

Attribute Preference and Discrimination Shifts in Young Children[1]

TOM TRABASSO, MARY STAVE, AND ROBERT EICHBERG
University of California, Los Angeles

The relations among preschool children's preferences for color and form attributes, speed of learning an initial problem, and optional shift behavior in discrimination training were studied in two experiments. In both experiments, learning was rapid if *S*'s preferred attribute was relevant. In Exp. I, upper-middle class *Ss* tended to reverse on their preferred attribute with planometric stimuli; in Exp. II, lower-middle class *Ss* who preferred color learned to discriminate stimulus objects slowly and made nonreversal responses regardless of the attribute upon which they were trained; form-preferring *Ss* showed the opposite behaviors. The results were discussed in terms of attentional and mediational theories.

Since Kendler and Kendler (1962) compared performance by rats, children, and adults on discrimination problems, considerable interest in the development of this problem-solving ability has evolved. While phylogenetic and developmental differences are frequent and fairly reliable, their theoretical interpretation has become a subject of controversy (Mackintosh, 1965; Kendler and Kendler, 1966; Wolff, 1967).

We may classify the alternative interpretations as mediational or attentional. To contrast these views, consider their respective interpretations of the Kendler and Kendler (1959) finding that kindergarten children who learned an initial discrimination slowly also learned a nonreversal more rapidly than a reversal problem; faster learners performed better on reversal shifts.

According to the mediational view espoused by the Kendlers, mature humans make nonovert mediating responses which act as conceptual labels and link different, related stimulus inputs and response outputs. For example, if red is the correct value, blue the wrong value, and triangles and circles are correct and wrong at random, the *S* learns to respond

[1]The authors thank Mrs. Dorothy Snyder, Director of the Children's Centers, Los Angeles City Schools, for her excellent cooperation in providing facilities and access to subjects for Experiment II. We also thank the University Elementary School of the University of California, Los Angeles, for similar aid in Experiment I. This research was supported by a research grant from the National Institutes of Health, United States Public Health Service, MH-08741.

© 1969 by Academic Press, Inc.

with the attribute label "color" to the red and blue values. Acquisition of a problem and a reversal shift is easier if one mediates since all that is required in learning is the association of the overt responses with the mediated values (e.g., in reversal, choose blue rather than red). A nonreversal shift, however, also requires the learning of a new mediating response (e.g., "form"). If one fails to mediate and responds to the specific stimulus values on the basis of their associative strength, then a nonreversal shift is easier; i.e., response strengths to triangles and circles are intermediate while those to blue are the weakest. To interpret the reported correlation between initial learning speed and rate of reversal, kindergarten children were assumed by the Kendlers to be in a transitional stage of mediational development in which some *Ss* (fast learners) are likely to mediate in the shift task while others (slow learners) are not.

Critics of mediational theory (cf. Wolff, 1967) have proposed that performance does not depend upon mediation or developmental ability but on whether the attribute to which the *S* is most likely to attend and use is relevant to the solution of the task. If *S* has a strong tendency to attend to an irrelevant attribute, he will learn a discrimination more slowly. In addition, assuming that the training is not sufficient to overcome the initial attention habit, he will do better on the nonreversal shift since his preferred attribute is now relevant. If, however, he initially attends to the relevant attribute, acquisition is more rapid; if one further assumes that the attention habit extinguishes more slowly than the instrumental response associations, reversal is also faster. These notions are readily deducible from a number of attentional models for discrimination learning, notably those of Zeaman and House (1963), Lovejoy (1968), and Trabasso and Bower (1968).

The assumption that learning is faster when *S*'s preferred attribute is relevant has been confirmed in several experiments (Heal, Bransky, and Mankinen, 1966; Smiley and Weir, 1966; Suchman and Trabasso, 1966b; Wolff, 1966). The interpretation that *Ss* reverse more readily on their preferred attribute has been supported in the studies by Heal *et al.* and Smiley and Weir. Subjects were trained with their preferred or nonpreferred attributes relevant, and made reversal or nonreversal responses to their preferred attribute. Both experiments thus replicated the Kendler's (1959) empirical relation between speed of initial learning and ease of reversal but account for the relation in terms of *S*'s attribute preference.

The present paper reports two experiments. Experiment I resembles the Smiley and Weir (1966) study in that an index of *S*'s preference for color or form stimuli was obtained independent of discrimination training and normal *Ss* were trained with their preferred or nonpreferred attribute relevant in an optional-shift paradigm. The main difference is that the

present experiment used a different method (the method of triads) to measure preference whereas both Smiley and Weir's and the Heal *et al.* experiments employed two-choice discrimination procedures throughout. However, since Exp. I was begun just prior to the publication of these studies, it should now be regarded as an attempt to replicate and generalize their findings.

Experiment II was designed to test whether or not the interaction between preference and learning was in terms of a preference for an attribute and not specific values. This was done by using identical attributes in the preference and training tasks but varying the attribute values. For example, suppose red and blue stimuli are used to test color preferences but *S*s are trained on green and yellow values. If preference is for color and this transfers in terms of attention to and use of different color values in training; i.e., *S* shows intradimensional transfer, he should learn quickly when color is relevant. Should this attention persist, *S* should also execute reversal response to color values on the optional-shift test.

EXPERIMENT I

Method

Stimulus Materials

The stimuli were drawings of triangles or circles (each approximately 1 sq. inch) colored either red or blue and presented on white, 4 × 6-inch file cards.

To measure preferences, a set of 18 cards was constructed. Each card contained three objects arranged in the pattern of an inverted triangle. For six cards, the color and form values were redundant (e.g., a card might contain two red triangles and one blue circle) so that a correct choice of a pair of objects could be based on either or both attributes. For the remaining 12 cards, the values were opposed (e.g., a card might contain a red triangle, a blue triangle, and a blue circle) so that a correct choice of a pair could be based on either color or form but not both. All combinations of values were used and their positions were counterbalanced over the sets.

For discrimination training, a set of four cards was used. Each card contained a pair of objects and the four cards represented all possible combinations and positions of the color and form values.

Procedure

The *S*s were tested and trained individually in a quiet room. Each *S* was seated at a small table opposite *E*. To begin, *S* was shown a preference card with redundant figures and heard the following instructions:

> "Here is a card. Each card has three pictures on it. Look at each card and tell me which two pictures are the same. Point to the two that are the same like this."

The *E* then pointed to the identical figures. Each of the remaining five redundant cards was presented successively and *S* self-paced his choices. If the choices were correct, *E* said, "Good, that is the idea." If *S* made an error, *E* said, "No, these two are alike," and pointed to the two identical figures.

After showing redundant set, *E* presented successively and without interruption the 12 cards that opposed color and form values. No feedback on choices was given during this series.

After the preference testing was completed, discrimination training began with *E* reading the following set of instructions:

> "O.K. You did very well. Now we are going to play a guessing game. Here is a card (chosen at random) with only two pictures on it. I am thinking very hard about one picture. Can you tell me which one?"

If a choice was correct, *E* reinforced the response by saying, "Good, that is the one I am thinking of" or "That is right." If a choice was wrong, *E* pointed to the correct figure and said, "No, I was thinking of this one." After receiving the appropriate feedback on his first choice, *S* was then presented one card at a time, self-paced his responses, and was given feedback. The intertrial interval was approximately 5 seconds.

The discrimination training and testing followed that of the optional-shift paradigm originated by Kendler, Kendler, and Learnard (1962). In Phase I, all four stimulus patterns were used and one attribute was relevant. After *S* met a criterion of 10 consecutive correct responses or fifty trials (whichever occurred first), he was shifted without interruption onto the second phase of training. In Phase III, a test phase of ten trials, all responses were reinforced as correct. Those *S*s who failed to meet the learning criterion in both training phases were not tested.

The attribute value and pattern combinations used in training and testing were completely counterbalanced over *S*s within each experimental condition; there were $4 \times 2 = 8$ such subgroups within each condition. The stimulus presentation orders for preference testing and each phase of discrimination learning were respectively random for each *S*.

Experimental Design

A *S* was classified as preferring color (or form) if 8 of 12 choices on the opposed-cue tests were for figures that were identical in color (or

form) and as "mixed" if neither criteria was met. Those *S*s who exhibited a preference were then randomly assigned to one of two conditions where either color or form was relevant. Thus, Preference (color or form) constituted one factor, and the Relevant Attribute of Phase I (color or form) constituted the second factor of a 2×2 design.

Subjects

The *S*s were 44 nursery and kindergarten children who attended the University Elementary School of the University of California, Los Angeles. Half of the *S*s were boys. The median age was 4 years, 8 months and the range in age was from 4 years, 2 months to 5 years, 4 months.

RESULTS

Attribute Preference

Twenty-eight, 15, and 1 *S*s were respectively classified as preferring form, color, or neither attribute. For those who had a preference, all but four made 11 or 12 consistent choices. The correlation between age and number of form choices was not significant ($r = -.04$). The latter result most likely arose from the restricted age range (14 months).

Discrimination Training

Eight and 7 color-preferring children and 14 form-preferring children each were trained with color or form relevant, respectively. Since errors and the trial of last error were nearly perfectly correlated ($r = .96$), analyses are reported only for errors. In Phase I, only the interaction between the preferred and relevant attributes was reliable ($F = 34.18$, $df = 1, 39$; $p < .01$). The mean number of errors for each of the four conditions of this interaction are given in Table 1.

In Table 1, it can be seen that ease of learning the initial discrimination depended upon the congruence between *S*'s preferred and relevant attributes. If his preferred attribute was relevant, fewer errors were made.

In Phase II, three *S*s were lost because of time limitations or refusals to continue and three more *S*s failed to learn. These six *S*s were not tested

TABLE 1
EXPERIMENT I: MEAN ERRORS IN PHASE I

Preferred attribute	Training attribute: Color	Training attribute: Form
Color	7.71	17.12
Form	19.79	2.28

in Phase III. The six comprised two color-preferring children in each of the training conditions and two form-preferring children who were trained on color. These six *Ss* were excluded from the Phase I results and a reanalysis on errors yielded the same reliable interaction ($F = 25.40$, $df = 1, 33$; $p < .01$).

The analysis of variance was performed on errors in Phase II. Again, only the interaction between preference and training attributes was reliable ($F = 4.94$, $df = 1, 33$; $p < .05$). The means for this interaction are given in Table 2.

TABLE 2
EXPERIMENT I: MEAN ERRORS IN PHASE II

Preferred attribute	Training attribute	
	Color	Form
Color	1.00	3.60
Form	2.25	1.21

In Table 2, *Ss* learned the redundant, reversal problem of Phase II more rapidly if they were trained with their preferred attribute relevant. Individual differences in ease of learning Phase I and Phase II were examined by correlating errors in the two phases; the correlation was positive ($r = .28$).

Optional Shifts: Phase III

The ages (in months) of the 37 *Ss* who learned Phase II and were tested in Phase III were analyzed as before and no significant factors or interactions were found. The correlation between Phase I errors and number of reversals in Phase III was $-.43$ ($p < .05$). Again, as in the learning of Phases I and III, only the preferred by trained attribute interaction was significant ($F = 4.00$, $p = .05$). The mean number of reversal

TABLE 3
EXPERIMENT I: MEAN NUMBER OF REVERSAL RESPONSES IN PHASE III

Preferred attribute	Trained attribute	
	Color	Form
Color	8.67	6.80
Form	3.75	7.43

responses by the groups contributing to this interaction is given in Table 3.

In Table 3, it can be seen that *Ss* tended to reverse on their preferred dimension. This relationship was also examined in terms of the number of *Ss* who met a criterion of eight or more reversal responses on the ten test trials of Phase III. These frequencies are summarized in detail in Table 4.

TABLE 4
EXPERIMENT I: NUMBER OF SUBJECTS MEETING VARIOUS REVERSAL CRITERIA IN PHASE III

Attribute		Number of reversal responses		
Preferred	Trained	8–10 (reversal)	0–2 (nonreversal)	3–7
Color	Color	5	1	0
Form	Form	10	2	2
Color	Form	2	0	3
Form	Color	4	7	1

The data in Table 4 were collapsed in various ways and subjected to chi-square analyses. It was found that individual *Ss* reversed more often on their preferred attribute ($X^2 = 4.77$, $df = 1$, $p < .05$), independent of which attribute was preferred ($X^2 = 1.93$) but more often when form was the training attribute ($X^2 = 3.77$).

Assuming that *S*'s preference for an attribute is an index of his initial tendency to attend to and use that attribute as the basis of his responding, the results of Exp. I support an attentional explanation of the empirical relationship between rate of initial discrimination learning and reversal shift behavior. The results indicated that ease of learning and rate of reversal were highly correlated, that ease of learning was also related to the function of *S*'s preferred attribute; namely, if it was relevant, learning was more rapid and that rate of reversal depended upon *S*'s attribute preference. However, it should be stressed that this interpretation of the present results also assumes that training effects do not offset the initial strengths of attending to one's preferred attribute versus one's nonpreferred attribute. It is possible in studies such as the present and that of Smiley and Weir (1966), which is essentially replicated by Exp. I, for *Ss* to reverse regardless of which attribute they are trained on if the training is of sufficient length to alter the initial, relative strengths of attending.

EXPERIMENT II

Method

Apparatus and Stimulus Materials

The stimuli for Exp. II were three-dimensional, wooden objects fashioned from 2-cu. inch blocks. Each block had a round hollow underneath for hiding trinkets. Two sets of colors and shapes were used: red and blue spheres and cubes; yellow and green tents and cylinders.

For discrimination training, the stimuli were displayed on a circular platform (15-inch diameter) which was perpendicularly bisected by a rectangle ($8 \times 15 \times \frac{1}{2}$ inches). The barrier and circle were mounted on a rotating base so that while *S* made choices between displayed objects on one side, *E* prepared the display on the other for the next trial. To *E*'s right was an 18×18-inch barrier used for storage of stimuli, data sheets, and reinforcements.

Procedure

The measurement of attribute preference and discrimination training were conducted by different *E*'s approximately one week apart. The assignment of *S*s to training conditions was random and was done double-blind. This procedure resulted in unequal numbers of *S*s in the experimental conditions but minimized possible *E* bias.

To measure a preference, three objects were placed on a low table about 3 inches apart in a triangular arrangement with the apex towards *S*. The instructions, order of redundant and opposed-attribute tests, appropriate feedback, method of triads, and pacing, were basically the same as those used in Exp. I. Each stimulus set was used to test one half of the *S*s. There were 12 redundant and 24 opposed-attribute arrangements, representing all combinations of objects and positions; the order of presentation was random for each *S*. The criteria for a preference was 18 or more choices for an attribute in the 24 opposed-attribute tests. Each *S* was promised a plastic ring of his choice, which he was given upon completing the preference testing.

All training procedures for Exp. II were identical to those of Exp. I with the following exceptions: (1) the stimuli were three-dimensional objects rather than two-dimensional drawings, (2) plastic trinkets were used as reinforcements in addition to verbal feedback, and (3) the learning criterion was ten successive correct responses or 30 trials. A *S* was allowed to keep three trinkets at the end of the experiment. One stimulus set was used throughout all training phases and half of the *S*s were assigned on a random basis to each set.

Experimental Design

The design of Exp. II was identical to that of Exp. I with one additional factor: the sets of stimuli used in preference testing and in discrimination training were either identical or different. The design was a $2^3 = 8$ factorial experiment with the following orthogonal factors: Preference (color or form), Training Attribute (color or form), and Stimulus Sets (same or different).

Subjects

The *S*s were 193 preschool children from Los Angeles City Schools' Children's Centers. Children in these centers are primarily from low-income families, headed by only one parent, who works, or by two parents, one working and the other either ill or attending school full time or working in an "essential industry" (e.g., nursing, teaching, defense plant). A sharply increasing fee schedule based on family income tends to discourage families of moderate-to-high income from enrolling their children in these centers.

There were 101 girls and 92 boys. The median age was 4 years, 7 months and the range was from 2 years, 8 months to 6 years. Eighteen *S*s failed to perform on the preference test; those *S*s whose preferences were measured had a median age of 4 years, 8 months, identical to that found in Exp. I. Twenty-two *S*s did not complete discrimination training because of absence from school.

RESULTS

References

The preference data were summarized in terms of the group percentage of color, form, or indeterminate choices (the chosen pair differed in both attributes) and in terms of the number of *S*s who were classified as having a reliable preference. These data for 6-month age groups are given in Table 5.

In Table 5 it can be seen that color was preferred more than form but that there was no clear developmental trend in preference. The correlations between age and color or form choices were —.01 and .02, respectively. The number of indeterminate and mixed preferences declined with age. If one excludes the youngest group, then the commonly reported age-form preference trend is observed. It should be noted that age is generally confounded by school location in this study so that the failure to show a marked developmental trend in form preference is not readily interpretable.

TABLE 5
EXPERIMENT II: PERCENTAGE OF GROUP CHOICES AND INDIVIDUAL PREFERENCES FOR COLOR AND FORM ATTRIBUTES

Age	Number *Ss*	Group			Individual		
		Color	Form	Indeterminate	Color	Form	Mixed
3.1–3.11	37	47	51	2	35	38	27
4.0–4.5	35	70	28	2	60	23	17
4.6–4.11	37	63	37	0	62	35	3
5.0–5.5	33	55	45	0	52	42	6
5.6–6.0	33	48	52	0	45	52	3
Total	175	57	42	1	51	38	11

Training

As found in Exp. I, errors and trial of last error were nearly perfectly correlated ($r = .94$) so that only results on errors are reported.

Of the 175 *Ss* who were tested for preference, 16 failed to show a reliable preference for color or form and an additional 22 did not undergo training. The errors made in Phase I by the remaining 137 *Ss* were analyzed in a $2 \times 2 \times 2$ analysis of variance with unequal n's. In this analysis, contrary to Exp. I, the only reliable factor was the preference ($F = 84.59$, $df = 1, 129$; $p < .01$). Children who preferred color learned considerably more slowly than those who preferred form; their respective mean errors were 12.52 and 3.66. This result is in line with the general finding that children who prefer color are poorer problem-solvers than those who prefer form (Suchman and Trabasso, 1966a).

The reliable interaction between preferred and trained attribute of Exp. I approached significance in Exp. II ($F = 3.73$, $p < .10$). The means for this interaction are given in Table 6.

TABLE 6
EXPERIMENT II: MEAN ERRORS IN PHASE I

Preferred attribute	Training attribute	
	Color	Form
Color	11.67	13.65
Form	4.39	2.50

In Table 6, *Ss* who were trained on their preferred attribute tended to learn more rapidly, a result in line with the attention position. The null result on different versus same attributes in preference testing and discrimination training indicates that these tendencies generalize to other

values of the preferred attribute, a result also in line with the attention hypothesis.

In Exp. I, *Ss* who did not learn Phase II were excluded from optional-shift testing in Phase III on the grounds that if *Ss* did not learn, then the tests would not be meaningful. This same procedure was used in Exp. II. As a result, 32 *Ss* did not undergo testing in Phase III. In the following, then, we report analyses on 105 *Ss* who underwent all three phases of discrimination training and testing. Table 7 gives the distribution and ages of these *Ss* in the eight experimental conditions; the number of *Ss* in parentheses is the number that failed to learn in Phase II.

TABLE 7

EXPERIMENT II: AGE DISTRIBUTION OF SUBJECTS WHO WERE TESTED IN PHASE III

Attribute		Stimulus sets			
		Same		Different	
Preferred	Trained	Age (months)	*N*	Age	*N*
Color	Color	59.12	17 (5)	56.64	14 (7)
Color	Form	56.09	11 (7)	58.20	10 (4)
Form	Color	57.36	14 (5)	57.47	17 (2)
Form	Form	57.78	9 (1)	54.85	13 (1)

The ages in Table 7 were also subjected to the analysis of variance and all resulting *F*-ratios were less than 1 in value. Thus, our randomization procedures were successful in matching groups on age and this factor does not confound interpretation of any findings. The analysis of variance on Phase I errors was repeated for the 105 *Ss* who completed all phases of training and testing. The results were the same as those reported above; namely, *Ss* who preferred color learned more slowly ($F = 86.36$, $df = 1, 97$; $p < .01$) while the interaction between preferred and trained attribute became reliable ($F = 4.95$, $df = 1, 97$; $p < .05$). Thus, excluding those *Ss* who failed to solve Phase II, the major finding of Exp. I was replicated in Exp. II.

The analysis of variance was performed on errors in Phase II. As in Exp. I, only the interaction between preferred and trained attributes was reliable ($F = 5.58$, $df = 1, 97$, $p < .05$). The means for this interaction are given in Table 8.

As can be seen in Table 8, *Ss* reversed faster when they were trained on their preferred attribute. Finally, errors in Phases I and II were correlated ($r = .39$, $p < .05$).

TABLE 8
EXPERIMENT II: MEAN ERRORS IN PHASE II

Preferred attribute	Training attribute	
	Color	Form
Color	3.61	4.87
Form	4.86	3.68

Optional Shifts

The initial learning rate-reversal relationship was examined by correlating errors in Phase I with the number of reversal responses in Phase III. This correlation was reliable and negative ($r = -.64$, $p < .01$), and is consistent with the Kendlers' (1959) finding.

The number of reversal responses in Phase III were analyzed in the $2 \times 2 \times 2$ analysis of variance. Contrary to the finding of Exp. I, the only reliable factor was that of Preference ($F = 48.34$, $df = 1, 97$; $p < .01$). Children who preferred color tended to make nonreversal responses. The mean number of reversal responses for *Ss* who preferred color was 1.73; for those who preferred form, 8.98. The *F*-ratio for the interaction between Preference and Trained Attribute was less than 1. The latter result would indicate that training offset initial tendencies to attend to one's preferred attribute.

The optional-shift data were analyzed in terms of the distribution of reversal responses made by individual *Ss* in the conditions. These data are summarized in detail in Table 9.

The data in Table 9 were collapsed and subjected to chi-square analyses. It was found that children who preferred color were classified more frequently as nonreversal *Ss* while those who preferred form were mostly

TABLE 9
EXPERIMENT II: NUMBER OF SUBJECTS MEETING VARIOUS REVERSAL CRITERIA IN PHASE III

Attribute		Number of reversal responses		
Preferred	Trained	8–10 Reversal	0–2 Nonreversal	3–7 Mixed
Color	Color	4	23	4
Form	Form	21	1	0
Color	Form	2	18	1
Form	Color	25	2	4

reversal *Ss* ($X^2 = 63.57$, $df = 1$, $p < .01$); there was no relationship between the kind of shift and either training on a preferred or nonpreferred attribute ($X^2 = .40$) or the attribute upon which *Ss* were trained ($X^2 = .01$).

DISCUSSION

Experiments I and II indicate that a *S*'s preference is related to his problem-solving ability and discrimination-reversal performance. In both experiments, there was an interaction between an attribute's function and degree of preference. If *S*'s preferred attribute was relevant, the initial discrimination and second phase reversal were more rapid. However, in Exp. II, those *Ss* who preferred color learned more slowly and failed to learn more frequently in both Phase I and II. In the third, optional-shift phase, *Ss* tended to reverse on their preferred attribute more often in Exp. I; in Exp. II, *Ss* who preferred form made more reversal choices while those who preferred color made mostly nonreversal choices, regardless of the attribute upon which they had been trained.

As far as attentional theory is concerned, these findings support, in part, alternative expectations that are derivable from such models. The results of Exp. I, for example, support the hypothesis that the training was not sufficient to offset the initial habits of attending to one's preferred attribute. Hence, *Ss* showed more reversal responses on their preferred attribute, independent of the attribute upon which they were trained. The performance of *Ss* who preferred form in Exp. II support the alternative attentional expectation; namely, that the training on the nonpreferred attribute was sufficient to overcome the initial tendency to respond to one's preferred attribute. Although learning and reversal were easier on relevant and preferred attributes, *Ss* who preferred form showed 90% reversal responses on the optional-shift tests, independent of the attribute upon which they were trained.

The results on the children who preferred color in Exp. II are inconsistent with either alternative attention hypothesis. Those *Ss* who had color as a relevant attribute learned and reversed faster than those who had form relevant. The fact that the latter *Ss* made more nonreversal responses in Phase III is consistent with the second attention hypothesis. However, what is inexplicable is the fact that nearly two-thirds of all color-preferring children who were trained on color made nonreversal responses; i.e., they chose the nonpreferred, randomly reinforced attribute.

This inconsistent finding as well as the one that the other color-preferring *Ss* made mostly nonreversal responses may be taken as favorable to a developmental theory of problem solving with some reservation

since age trends for preference were absent. If children who prefer color are considered to be less mature in their problem-solving ability (Suchman and Trabasso, 1966a), then the data of Exp. II support the Kendlers' (1962) position that problem-solving ability and reversal behavior are correlates of conceptual development.

A final point should be made about the differences in results for the two experiments on the optional-shift data. In general, one might regard Exp. II as the more sensitive experiment since it involved a larger number of *S*s, was run under a double-blind procedure which reduced the possibility of *E* bias, had a larger range of stimulus materials, and contained more *S*s who preferred color than either Exp. I or Smiley and Weir's (1966) experiment. There is the further difference of kind of *S*s sampled. The *S*s in Exp. I were primarily upper-middle class and experimentally sophisticated while those in Exp. II were largely lower-middle class and unsophisticated in psychological testing. It is difficult to assess what effect this kind of variable had in the present study but forewarns of relying exclusively upon convenient, university elementary school populations as sources of *S*s.

REFERENCES

Heal, L. W., Bransky, M. C., and Mankinen, R. L. The role of dimension preference in reversal and nonreversal shifts of retardates. *Psychonomic Science*, 1966, **6**, 509–510.

Kendler, H. H., and Kendler, T. S. Vertical and horizontal processes in problem solving. *Psychological Review*, 1962, **69**, 1–16.

Kendler, H. H., and Kendler, T. S. Selective attention versus mediation: Some comments on Mackintosh's analysis of two-stage models of discrimination learning. *Psychological Bulletin*, 1966, **66**, 282–288.

Kendler, T. S., and Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *Journal of Experimental Psychology*, 1959, **58**, 55–60.

Kendler, T. S., Kendler, H. H., and Learnard, B. Mediated responses to size and brightness as a function of age. *American Journal of Psychology*, 1962, **75**, 571–586.

Lovejoy, E. *Attention in discrimination learning*. San Francisco: Holden-Day, 1968.

Mackintosh, N. J. Selective attention in animal discrimination learning. *Psychological Bulletin*, 1965, **64**, 124–150.

Smiley, S. S., and Weir, M. W. The role of dimensional dominance in reversal and nonreversal shift behavior. *Journal of Experimental Child Psychology*, 1966, **4**, 296–307.

Suchman, R. G., and Trabasso, T. Color and form preference in young children. *Journal of Experimental Child Psychology*, 1966a, **3**, 177–187.

Suchman, R. G., and Trabasso, T. Stimulus preference and cue function in young children's concept attainment. *Journal of Experimental Child Psychology*, 1966b, **3**, 188–198.

Trabasso, T., and Bower, G. H. *Attention in learning*. New York: Wiley, 1968.

WOLFF, J. L. [illegible] role of dimensional preferences in discrimination learning. *Psychonomic Science*, 1966, **5**, 455–456.

WOLFF, J. L. Concept-shift and discrimination-reversal learning in humans. *Psychological Bulletin*, 1967, **68**, 369–408.

ZEAMAN, D., AND HOUSE, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963, pp. 159–223.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 210–221 (1969)

Age and Sex Differences in Responses to Embedded Figures and Reversible Figures[1]

LUDWIG IMMERGLUCK[2]
San Francisco State College

AND

MARIA CLAUDIA MEARINI[2]
Harvard Florence Research Project, Florence, Italy

Responses to Embedded Figures and to a series of Reversible Figures, as given by 9-year-old, 11-year-old, and 13-year-old subjects, were assessed with a view of identifying systematic age and sex differences. The present findings show that: (1) the ontogenetic course of sex differences in perceptual responses follows a zigzagging and overlapping pattern rather than a unidirectional path, and data obtained from adult subjects may offer no valid clues with regard to the existence or direction of sex differences at younger age levels; (2) a particular pattern of sex differences uncovered in a specific perceptual response category supplies no valid indices for the existence of similar sex difference patterns in other and even related response categories; and (3) age changes in perceptual responses, like those pertaining to the pattern of sex differences, also appear to be task, response, and in some cases, sex specific.

Although the existence of sex differences in perceptual responses has been repeatedly demonstrated in a wide array of investigations and also by researchers working in often diverse cultural settings (Andrieux, 1955, Bennett, 1956, Frank, 1956, Newbigging, 1954), a key question remains, apart from the obvious one regarding the causes of such

[1] This study was supported through a U. S. Public Health Service Special Fellowship award (1F3 MH-19-555) to Ludwig Immergluck for his work at the Harvard Florence Research Project in Florence, Italy.

[2] Grateful acknowledgment is made to the staff of the Harvard Florence Research Project, and especially to Dr. Boutourline Young for placing the institute's research resources at our disposal, and to Dr. Gino Tesi for his aid in preparing the test material, translation of the instructions, and statistical evaluation of the data. We are greatly indebted to Provveditore agli studi Prof. Silvio Peluso; Dr. Marcello Romanelli; Prof. Achille Guerra, inspector of elementary schools, and Prof. Centi, principal of the Dante Elementary School, Florence, Italy, for their invaluable help in securing the subject groups.

differences, with respect to the precise age level at which sex-linked differences emerge and the character of their ontogenetic course. Developmental studies addressed to this specific issue are surprisingly scarce, and, at any rate, even in otherwise rigorous ontogenetic investigations there is sometimes only a casual concern in evidence for the need to control for the sex of the subject groups, with an implicit assumption that the scrutinized age levels alone constitute the overriding factors which determine potential performance differences. This assumption may, of course, be perfectly justified in many instances, but in the light of our present-day knowledge of sex-linked behavioral differences, it would be most precarious to maintain any *a priori* presumptions regarding the potential relevance of such differences.

Any exploration aimed at pinpointing a specific age level at which sex differences in perceptual responses first emerge must by necessity also attend to some closely related issues. That is, once a particular age point has been clearly identified at which sex differences are first demonstrable, the question arises with regard to the maintenance of these differences along an ascending developmental continuum, and also whether the direction of any uncovered differences remains constant ontogenetically, reaching essentially unchanged into maturity. In adult *S*s, for example, it has been demonstrated consistently (Witkin, 1949a, b; Witkin *et al.*, 1954) that men achieve significantly higher "field-independent" scores on the Rod and Frame than do women, and similarly, male *S*s are significantly more skillful in perceptually uncovering a hidden geometric design from its obscuring context than are female *S*s.

But are the discovered sex differences in the particular perceptual response classes also demonstrable ontogenetically, or might there be particular moments in chronological development in which such differences are (1) not exhibited, (2) similar to those found in adult *S*s, (3) opposite to those found in adult *S*s, and, finally (4) can generalizations be drawn from one specific set of responses to other related response classes or are both age and sex differences task-specific?

It is to these general questions that the study at hand is addressed. More specifically, a comparison is sought between the responses to two particular types of perceptual tasks, an Embedded Figures task and a series of Reversible Figures, as they are given by boys and girls at different age levels: 9 years, 11 years, and 13 years.

The choice of the stimulus material was based upon a number of considerations. The Embedded Figures task was selected primarily because of its wide use in related perceptual research and its demonstrated relationship to a gamut of broader and more complex personality-related behavior patterns (e.g., Witkin, 1954, 1962). The Reversible Figures were

included because they represent an essentially different kind of stimulus material, calling on perceptual-response operations that are basically different from those involved in Embedded Figures, and also because a series of recent investigations (Immergluck, 1966a,b,c) revealed significant and consistent individual differences in reversal amplitude which were found to be systematically related to other perceptual response and behavior categories.

The particular age levels, 9, 11, and 13 years, were selected because it is at these phases within the developmental continuum that broad physiological, maturational, and behavioral changes occur, and more particularly changes in which significant sex differences might be logically anticipated to exist.

METHOD

Subjects. Three *S* groups composed of 20 boys and 20 girls each and representing the 9-, 11-, and 13-year age levels respectively (with a mean age of 9.4, 11.4, and 13.6 years) were selected from the primary and secondary public school system in Florence, Italy. All 120 *S*s (40 within each age group) came from a very similar middle-class socioeconomic background and represented average academic achievers for each school class.

Embedded Figures task. Each subject, in individual sessions, was presented with 15 pairs of the Gottschaldt Embedded Figures (Gottschaldt, 1926) printed on three sheets containing five designs each and preceded by a cover sheet containing printed instructions which were repeated and read orally by the *E*, and one pair of sample figures demonstrating and clarifying the nature of this particular task. After the read and orally presented instructions together with the sample figures, the *S* was then required to trace with a specially supplied colored pencil the simple geometric design figure appearing on the left side in the complex design context pattern on the right. The instructions referred to a 5-minute time limit for this task and each *S* was asked to work as quickly as possible. The *S*'s score is represented by the number of correct tracings within a 5-minute time module.

The Reversible Figures tasks. Three widely known ambiguous figures, Rubin's "Profile-Vase," Rubin's "Cross" (an octagon comprised of a black and white "windmill" cross), and the commonly employed reversible perspective "Prism" (a small square within a large square connected at their corners by diagonal lines) were respectively utilized as the Reversible Figures tasks. Each figure was presented for a 1-minute interval during which the *S*'s number of reversals (indicated orally by the *S*) was recorded.

RESULTS

Embedded Figures tasks. Table 1 compares the performance of boys and girls at each of the three age levels. At age 9 the female group achieved a significantly higher score ($\chi^2 = 5.01$, $p < .05$) than did the boys, but no significant differences were found at ages 11 and 13. Figure 1 compares the mean scores for the combined total *S* group as well as for the male and female *S* groups between each of the age spans (age 9 to 11: $\chi^2 = 31.42$; $p < .001$; age 11 to 13: $\chi^2 = 18.62$; $p < .001$).

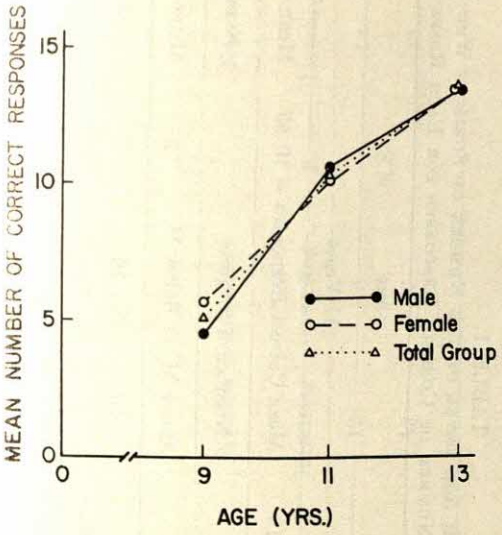


FIG. 1. Embedded Figures. Age and sex comparison of the mean number of correct responses.

Reversible Figures tasks. Table 2 contrasts the male and female *S* groups at each age level in terms of the number of *S*s who gave reversal responses to any or all reversible figures. Using this criterion, significant sex differences were found at only the 9-year level, at which a greater number of male *S*s gave reversal responses ($\chi^2 = 9.23$, $p < .01$).

An age level comparison of the total number of *S*s (combining male and female *S*s) who gave reversal responses shows (Table 3) a significant rise between age levels 9 and 11 ($\chi^2 = 7.2$, $p < .01$) but no significant difference between ages 11 and 13. Interestingly enough, a breakdown of the total *S* groups into male and female *S*s. (cf. Table 3 and Fig. 2) reveals that this significant age-related rise pertains only to the female *S*s ($\chi^2 = 12.9$, $p < .001$), while the number of *S*s showing reversal at these diverse age levels remains essentially unchanged for the male group.

TABLE 1

EMBEDDED FIGURES: A SEX COMPARISON AT EACH AGE LEVEL OF THE NUMBER OF SUBJECTS WHO GAVE CORRECT RESPONSES ABOVE AND BELOW THE MEAN NUMBER OF CORRECT RESPONSES FOR EACH RESPECTIVE AGE

	Age								
	9 Years			11 Years			13 Years		
	Mean Correct Responses = 5.15			Mean Correct Responses = 10.55			Mean Correct Responses = 13.57		
	Number of subjects			Number of subjects			Number of subjects		
Subjects	Above M	Below M	p	Above M	Below M	p	Above M	Below M	p
Male ($N = 20$)	5	15	$< .05$	10	10	N.S.	12	8	N.S.
Female ($N = 20$)	12	8		11	9		13	7	

TABLE 2

REVERSIBLE FIGURES: A SEX COMPARISON AT EACH AGE LEVEL OF THE NUMBER OF SUBJECTS GIVING REVERSAL RESPONSES

	Age								
	9 Years			11 Years			13 Years		
	Number of subjects			Number of subjects			Number of subjects		
Subjects	Reversing	Not reversing	p	Reversing	Not reversing	p	Reversing	Not reversing	p
Male ($N = 20$)	11	9	$< .01$	12	8	N.S.	12	8	N.S.
Female ($N = 20$)	2	18		13	7		12	8	

TABLE 3

REVERSIBLE FIGURES: AN AGE COMPARISON (BETWEEN 9 AND 11 YEARS AND 11 AND 13 YEARS) OF THE NUMBER OF SUBJECTS WHO GAVE REVERSAL RESPONSES

	Subjects										
	Total group				Male				Female			
		Number of subjects				Number of subjects				Number of subjects		
Age (years)	N	Reversing	Not reversing	p	N	Reversing	Not reversing	p	N	Reversing	Not reversing	p
9	(40)	13	27		(20)	11	9		(20)	2	18	
				$< .01$				N.S.				$< .01$
11	(40)	25	15		(20)	12	8		(20)	13	7	
11	(40)	25	15		(20)	12	8		(20)	13	7	
				N.S.				N.S.				N.S.
13	(40)	24	16		(20)	12	8		(20)	12	8	

Different results are yielded, both in terms of age and sex contrasts, through comparison of reversal rate for those *Ss* who gave reversal responses. Table 4 shows significant sex differences in reversal rate at each age level. That is, although the number of *Ss* who showed any reversal at all is essentially the same for both male and female *Ss* at ages 11 and 13, male *Ss* showed a significantly higher reversal rate at all age levels than did the female *Ss* (age 9: $\chi^2 = 4.6$, $p < .05$; age 11: $\chi^2 = 4.1$, $p < .05$; age 13: $\chi^2 = 4.4$, $p < .05$).

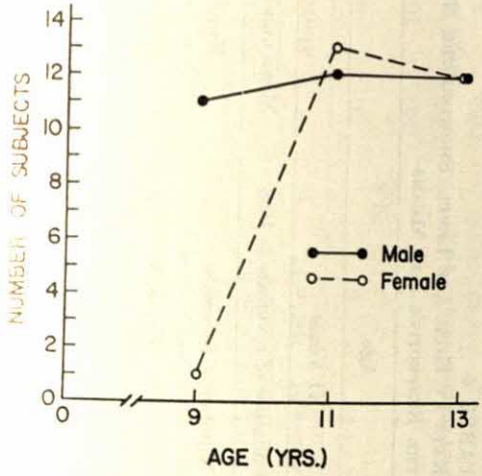


FIG. 2. Reversible Figures. Age and sex comparison of the number of subjects giving reversal responses.

No significant age-related rise in reversal rate is revealed when one considers total *S* group performance at each age level (Table 5) but separating these performance scores by sex shows the absence of age differences to hold only for male *Ss* while the rate of reversal increases significantly from age level to age level for the female *Ss* (age 9 to 11: $\chi^2 = 3.95$, $p < .05$; age 11 to 13: $\chi^2 = 4.8$, $p < .05$).

CONCLUSIONS

The presently obtained results pertaining to age and sex differences in responses given to a series of stipulated perceptual tasks suggest the following conclusions:

Data obtained from adult *Ss* may supply no valid clues with regard to the existence or direction of sex differences at younger age levels. In contrast to the reported sex differences in responses to Embedded Figures found in adult *Ss* both in the United States and abroad (Witkin, 1954; Andrieux, 1955; Newbigging, 1954), the presently obtained results reveal a female superiority at the 9-year age level and no significant sex dif-

TABLE 4

REVERSIBLE FIGURES: A SEX COMPARISON OF REVERSAL RATE AT EACH AGE LEVEL, SHOWING THE NUMBER OF REVERSALS ABOVE AND BELOW THE RESPECTIVE AGE MEANS

	Age											
	9 Years				11 Years				13 Years			
	Mean number of reversals = 8.1				Mean number of reversals = 10.8				Mean number of reversals = 16.4			
		Number of reversals				Number of reversals				Number of reversals		
Subjects	*N*	Above M	Below M	*p*	*N*	Above M	Below M	*p*	*N*	Above M	Below M	*p*
Male	(10)	7	3	$< .05$	(19)	11	8	$< .05$	(20)	13	7	$< .05$
Female	(3)	0	3		(16)	4	12		(19)	6	13	

TABLE 5

REVERSIBLE FIGURES: AN AGE COMPARISON (BETWEEN 9 AND 11 YEARS AND 11 AND 13 YEARS) OF REVERSAL RATE SHOWING THE NUMBER OF REVERSALS ABOVE AND BELOW THE RESPECTIVE COMBINED AGE MEANS

	Subjects											
	Total group				Males				Females			
	Mean reversal rate = 13.8				Mean reversal rate = 11.17				Mean reversal rate = 8.6			
Age (years)	N	Number of reversals: Above M	Number of reversals: Below M	p	N	Number of reversals: Above M	Number of reversals: Below M	p	N	Number of reversals: Above M	Number of reversals: Below M	p
9	(13)	3	10	N.S.	(10)	4	6	N.S.	(3)	0	3	$<.05$
11	(35)	10	25		(19)	7	12		(16)	10	6	
	Mean reversal rate = 13.7				Mean reversal rate = 14.33				Mean reversal rate = 11.9			
11	(35)	10	25	N.S.	(19)	6	13	N.S.	(16)	5	11	$<.05$
13	(39)	19	20		(20)	11	9		(19)	13	6	

ference at ages 11 and 13. The implication is strong that the ontogenetic course of sex differences in this type of perceptual response follows a zig-zagging and overlapping pattern rather than a unidirectional path.

The particular pattern of sex differences uncovered by ontogenetic studies in a stipulated response category offers no valid clues for the existence of similar sex difference patterns in other response categories, such as perception, even if these are within the same general behavioral class. In contrast to the responses given to the Embedded Figures tasks, in reversal rate on the Reversible Figures Tasks male *Ss* were clearly superior to the females at all of the scrutinized age levels. Sex differences, then, as far as ontogenetic development is concerned, appear to be task and response specific.

This finding is particularly significant in the light of recent studies (Immergluck, 1966a,b,c) which in adult *Ss* demonstrated a close relationship between reversal amplitude in Reversible Figures and perceptual measures of field-dependence (i.e., field-independent *Ss* showed a significantly higher reversal rate than did field-dependent *Ss*). Since responses to Embedded Figures are commonly taken as an index of a *S*'s field-dependence characteristic (correlating highly with other field-dependence measures), it may be logically assumed that responses to Embedded Figures and Reversible Figures should go together. The fact that this does not hold at all age levels—as the present data indicate—suggests that both the mechanisms and the psychological meaning underlying these discrete perceptual response classes might be different at varying age levels. In short, one can not be certain that a given set of perceptual responses which has a demonstrated usefulness as a valid measure of a given psychological dimension, such as field-dependence, in adult *Ss* provides a similarly valid index of such a psychological dimension in an ontogenetic context.

Age changes in perceptual responses, like those pertaining to the pattern of sex differences, appear also to be task and response specific, and in some instances sex specific. While performance on the Embedded Figures increased very sharply from lowest to highest age level, no such performance increases were obtained with the Reversible Figures. Separate examination, however, of performance scores for male and female *Ss* reveals no age-related increase in reversal amplitude for males, but consistent increases for female *Ss*. Similarly, the absolute number of *Ss* giving reversal responses at each age level remains quite unchanged for boys, but increases significantly for female *Ss*.

In an overview of the data at hand, it needs to be emphasized that the present study has tapped only a relatively narrow chronological age span, and, at that, employing but two circumscribed types of perceptual

tasks. But even this narrow age and stimulus range yields data which underscore both the complexity and specificity of ontogenetic perceptual development for each sex.

Finally, considering the fact that the present *S* groups were obtained in Italy, the issue of potential cultural differences, while none have been reported in an array of relevant studies employing similar stimulus designs, needs to be probed into in a more punctilious and systematic manner, particularly since such differences may have varying magnitude at diverse age levels.

REFERENCES

ANDRIEUX, C. Contribution a l'etude des differences entre hommes et femmes dans la perception spatiale. *L'anne Psychologique,* 1955, **55,** 41–60.

BENNETT, D. H. Perception of the upright in relation to body image. *Journal of Mental Sci...* 1956, **102,** 487–506.

FRANKS, C. M. Différences determinées par la personalité dans la perception visuelle de la verticale. *Revue de Psychologie Appliquée,* 1956, **6,** 235–246.

GOTTSCHALDT, K. Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren; I: Über den Einfluss gehaufter Einprägung von Figuren auf ihre Sichtbarkeit in umfassender Konfigurationen. *Psychologische Forschungen,* 1926, **8,** 261–317.

IMMERGLUCK, L. Visual figural after-effects and field dependence. *Psychonomic Science,* 1966a, **4,** 219–220.

IMMERGLUCK, L. Figural after-effects, rate of "figure-ground" reversal, and field dependence. *Psychonomic Science,* 1966b, **6,** 45–46.

IMMERGLUCK, L. Resistance to an optical illusion, figural after-effects, and field dependence. *Psychonomic Science,* 1966c, **6,** 281–282.

NEWBIGGING, P. L. The relationship between reversible perspective and embedded figures. *Canadian Journal of Psychology,* 1954, **8,** 204–208.

WITKIN, H. A. Sex differences in perception. *Transaction of the New York Academy of Science,* 1949a, **12,** 22–26.

WITKIN, H. A. The nature and importance of individual differences in perception, *Journal of Personality,* 1949b, **18,** 145–170.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., AND WAPNER, S. *Personality through Perception.* New York: Harper, 1954.

WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., AND KARP, S. A. *Psychological Differentiation.* New York: Wiley, 1962.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 222–233 (1969)

Maintenance of Children's Behavior by Accidental Schedules of Reinforcement[1]

PAUL WEISBERG AND DALE B. KENNEDY
University of Alabama and Rhode Island University

Ss were trained to lever press for snack reinforcements under response-dependent schedules—either a variable interval 30 seconds, limited hold, 10 seconds contingency (Exp. 1), or a variable ratio 15 contingency (Exp. 2). Following these baselines the schedule conditions were changed and most *Ss* were placed on extinction or on response-independent (accidental) reinforcement schedules. Accidental reinforcement in Exp. 1 was dispensed for all *Ss* at variable intervals having as an average, 30 seconds (Acc. VI 30 seconds) and in Exp. 2, dispensed at fixed intervals (Acc. FI), the duration of each interval not the same for every *S*, but for most was near 16 seconds. In both Exp. 1 and 2 the accidental schedules maintained the baseline response rates at higher levels and for more sessions than did extinction operations. The rates of proportionally more *Ss* were sustained by the Acc. FI (Exp. 2) than by the Acc. VI 30 seconds (Exp. 1); three Exp.-2 *Ss* after seven Acc. FI sessions were still responding at levels consistent with or higher than baseline. In Exp. 1 and 2, the correlation between baseline and accidental reinforcement rates was greater than that found between baseline and extinction.

The experimental arrangement whereby a reinforcing agent is presented noncontingent upon behavior is called an accidental schedule of reinforcement. As originally demonstrated by Skinner (1948) and more recently by Herrnstein (1966), even though the contingency between some behavior and a known reinforcer is not deliberately programmed, a capricious relationship may nonetheless form and endure almost indefinitely. Accidental or noncontingent schedules may have two important functions: They may foster the development of new and "superstitious" activities and/or they may maintain some already acquired behavioral pattern.

In regard to the contribution of accidental schedules to the acquisition of behavior, many investigators (cf. Church, 1964) have administered

[1]This research was supported by a PHS grant (MH 13877-01) awarded to the first author while he was at Brown University. The authors are grateful to Mrs. Jane E. Walsh and the staff of Lakeside Shelter, Warwick, R. I., and to Sister Mary Paula of Carter Day Nursery, Providence, R. I.

response-independent reinforcing events to a separate (yoked) control group as a means of testing possible "sensitization" and "activation" effects caused by the reinforcing events themselves. In general the results of these studies show that the delivery of reinforcers noncontingent upon behavior at the most only moderately enchances the to-be-conditioned response. Although the yoked control design may bias the frequency of response in favor of an experimental group receiving response-contingent events (Church, 1964), delay of reinforcement gradient studies (Kimble, 1961) provide independent support for the contention that rapid rates of acquisition predominate when the consequences are systematically and immediately applied following some specified behavior.

Methodologically, it seems easier to study the behavior-maintenance functions of accidental reinforcement schedules than to observe how it produces new and unknown forms of behavior. Thus a major difficulty in charting the development of superstitious sequences is the inability to specify ahead of time the ultimate form that these idiosyncratic behaviors will take. A classification scheme to measure continuously and reliably all the behaviors which occur when an organism is given reinforcements *ad lib* may prove infeasible, unless one has access to sophisticated video and recording equipment.

The present experiments considered the second function of accidental reinforcement schedules: To what extent and under what conditions can response-independent events sustain a child's behavior that was previously conditioned by response-dependent schedules of reinforcement? Experimentally, this question can easily be implemented and cause little in the way of behavioral recording problems since a response must be identified and brought under the control of deliberate contingencies before the effects of accidental schedules are assessed.

METHOD

Subject and setting. Experiment I *S*s were in temporary residence at a shelter home while Exp. 2 *S*s attended a nursery school. Socio-economic level of the parents of both groups ranged from lower to upper lower class. Age of each *S* (in months) can be located in Figures 1, 2, and 3. Throughout testing sessions, *S* sat strapped in a highchair before a large wooden superstructure, the inside of which contained a 70-bucket Universal feeder and mounted to the exterior of which were a food tray and a red plastic lever. The lever was on *S*'s right side and when pressed, it produced an audible click. About 6 feet from the rear of the superstructure was an equipment cart, housing a solid-state programming

console, impulse counters and a cumulative recorder. An adult observer (*O*), seated behind the cart and mostly obscured from *S*'s view, monitored the console.

Reinforcers consisted of a variety of snacks broken into small pieces (cinnamon crackers, marshmallows, chocolate chips, potato sticks). Daily snacks, usually given as part of "snack time" were omitted, and *S*s were tested from 2½ to 3 hours after their last meal.

General procedure. Following a familiarization period, in which *S*s were encouraged to accept snacks first from *O*'s hand and then from the food tray, *S*s were individually magazine trained (Weisberg and Fink, 1966; Weisberg and Tragakis, 1967). *O*, seated next to *S*, dispensed snacks via a microswitch connected to the feeder (the lever was removed), and took care not to reinforce the occurrence of superstitious behavior. *O* then gradually increased the number of long (over 1 minute) inter-reinforcement intervals in order to build a tolerance to the display of emotional behavior. Gradually, social and verbal encouragement were withdrawn altogether and the spatial distance between *O* and *S* increased successively until *O* delivered snacks while stationed behind the equipment cart. Since these snacks were intended to constitute the major source or reinforcement control, henceforth, during the entire duration of each session, *O* assumed a perfunctory role, allowing only minimal verbal and social interaction.

During subsequent acquisition and rate-strengthening sessions the lever was available and presses were reinforced over several fixed ratio values in the direction of an FR-10 schedule. Under the FR 10, all *S*s received at least 30 reinforcements within one session and emitted at least 20 responses per minute (rpm). In case an *S* failed to respond within 3 minutes of the first acquisition session, *O* demonstrated, without verbally instructing, by passively placing *S*'s hand on the lever until three to five snacks were earned; this "priming" procedure was necessary only for *S*4, *S*11, and *S*16. The rate-strengthening procedure together with the magazine training was usually accomplished within four sessions.

Four *S*s from Exp. 1 and three *S*s from Exp. 2 were dropped because of failure to establish high (at least 20 rpm) and stable rates. Other *S*s did not finish all the experimental conditions planned for them; an account of and explanation for these incompletions will be provided in other sections of this paper. *S*s were run at least 5 days a week, breaks being caused by temporary illness (*S*6, *S*17) or when an *S* was briefly united with his parents (*S*1). No new condition was instituted following a break of 2 or more days between adjacent sessions.

Experiment 1

In the session following FR-10 experience, lever pressing was reinforced according to a VI 30-second program tape having a range of 1–120 seconds (Flesher and Hoffman, 1962). A scheduled reinforcer remained available only for 10 seconds, after which failure to respond advanced the programming tape and erased the reinforcer. This "limited hold" (LH) 10-second contingency, by penalizing unduly long pauses, reduced their frequency and tended to produce steady responding throughout the VI 30 seconds. The criterion adopted for the maintenance of conditioned behavior was to run *Ss* for at least five sessions (40 reinforcements per session) under the VI 30-second, LH 10-seconds schedule until a rate of 20 rpm or greater was established for two consecutive sessions.

Nine *Ss* fulfilled these baseline requirements. During the next experimental phase, the programming and delivery of reinforcers were altered: Six *Ss* were placed on an accidental or noncontingent reinforcement schedule, whereby reinforcements (40 per session) were delivered according to the previously used VI 30-second tape but they were given in the independence of *Ss* responding (i.e., Acc. VI 30-second schedule) and the other three *Ss* were not reinforced for responding (extinction), the VI 30-second tape used solely to define the duration of a session. Accidental reinforcement and extinction conditions remained in force for several sessions until each *S*'s rate of responding dropped and remained below 50% of the value established by the last two VI 30-second, LH 10-second baseline sessions.

Following accidental reinforcement training for the six *Ss* on this regime, baseline recovery was attempted; i.e., rate-strengthening by FR schedules and reinstatement of the VI-30 second, LH 10-second contingencies. Four of these Ss were then switched to extinction. (Transfer of two children to their real or foster home accounted for subject loss.)

RESULTS

Figures 1 and 2 relate the changes in each *S*'s response rate to the various modifications in the reinforcement contingency. Only the rates of the last three sessions of each *S*'s original baseline (Figures 1A and 2A) and of baseline recovery (Figure 1C) are shown. Omitted, therefore, are the rates stemming from the rate strengthening sessions and the four to five baselines sessions which usually preceded the last three presented. Comparison of Figure 1B and 2B indicates that the operations of noncontingent reinforcement and nonreinforcement func-

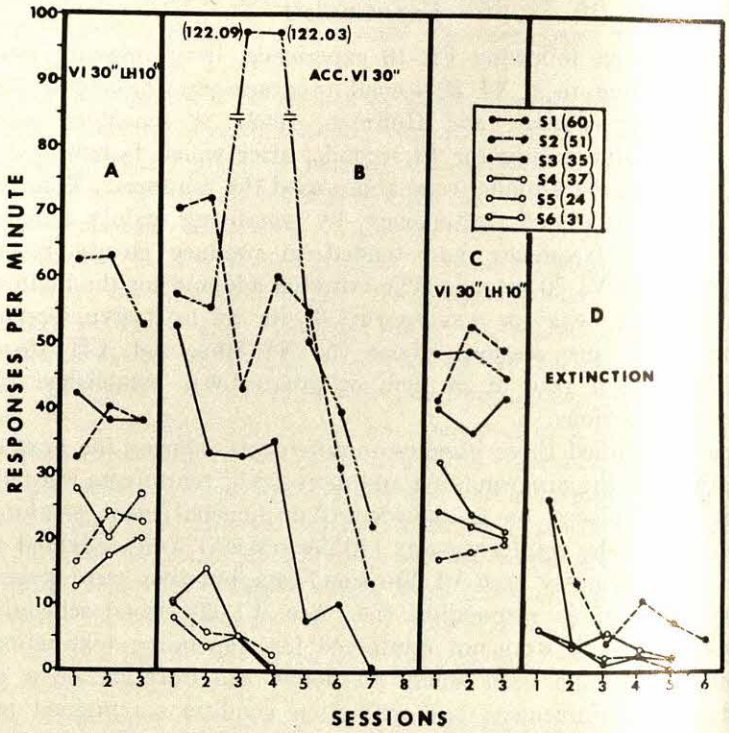


FIG. 1. Changes in response rate over experimental conditions. Only the last three sessions of the original VI 30-second, LH 10-second baseline and the recovery baseline are shown.

tioned to reduce previous conditioned behavior. However, the proportion of *S*s who showed rapid reduction as against a gradual decline in response rate differed. Extinction had the effect over time of abruptly lowering response rate for all three *S*s while the rate of decline for three (*S*1, *S*2, and *S*3) of the six *S*s placed on accidental reinforcement was slower. The extinction data in Figure 1D of four *S*s (*S*2, *S*3, *S*5, and *S*6) who were placed on extinction following a second baseline test also suggests that the process produced by extinction and the Acc. VI 30-second schedule are not the same. However, since it is common for behavior to progressively weaken upon repeated extinction tests (cf. Millenson, 1967), this last finding may be equally interpreted as an order effect.

References to Figures 1A and 1B shows that the degree to which the noncontingent schedule sustained behavior dependent somewhat upon the level of baseline responding with *S*1, *S*2, and *S*3 having the highest

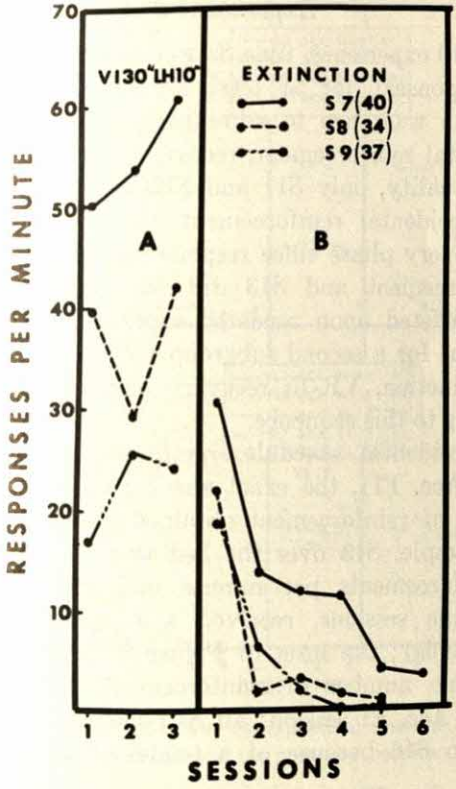


FIG. 2. Extinction performance following the last three baseline sessions.

baseline rates and being maintained longest by the Acc. VI 30-second schedule. On the other hand, the relationship between the level of baseline responding and the fall in response across extinction sessions is less clear (contrast the intra-subject functions in Figures 1C and 1D, and 2A and 2B). Since the programming of response-independent events exerted weak control for *S*s having low conditioning rates (*S*4, *S*5, and *S*6 in Figure 1A), Exp. 2 was conducted to enhance the influence of accidental contingencies by giving *S*s a conditioning history which would guarantee a higher rate of responding. In Exp. 2, reinforcement during baseline was made contingent upon the fulfillment of a response criterion by the utilization of a variable ratio 15 schedule, a schedule found to yield higher rates than a VI 30-second schedule (Weisberg, 1969). Also in Exp. 2, the noncontingent reinforcement-extinction sequence was counterbalanced and, during accidental reinforcement sessions, snacks were presented at fixed rather than at variable intervals.

Experiment 2

Following FR-10 experience, nine *S*s were placed on a VR-15 schedule (range, 4–40 responses) for at least six sessions. For one subgroup ($N = 4$), the plan was then to administer, a number of sessions consisting of accidental reinforcement, recovery of the VR-15 baseline, and extinction. In actuality, only *S*11 and *S*12 kept to this order; *S*10 was switched from accidental reinforcement directly into extinction, skipping the VR recovery phase since responding did not deteriorate during accidental reinforcement and *S*13 did not complete extinction since he cried and protested upon repeated exposure to this condition. The order of conditions for a second subgroup ($N = 5$), following the VR-15 baseline was extinction, VR-15 recovery and accidental reinforcement with all *S*s keeping to this sequence.

During the accidental schedule, reinforcements were presented at fixed intervals (Acc. FI), the exact spacing of which was based upon the average rate of reinforcement obtained under the last two VR-15 sessions. For example, *S*13 over the last two VR-15 sessions averaged about four reinforcements per minute and therefore throughout all accidental schedule sessions, received a reinforcer every 15 seconds (Acc. FI 15 seconds); the inset of Figure 3 specifies the FI value for each *S*. The same number of reinforcements were dispensed during each VR 15 and Acc. FI session; all *S*s received 35 snacks except that 25 were given to *S*16 because of a tendency to satiate early in the session.

RESULTS

Figure 3 plots the response rates to the two treatment orders, the top panel illustrating the process when the Acc. FI schedule preceded extinction and the bottom panel when it followed extinction. Again, only the rates of the last three VR-15 baseline sessions are displayed.

As in Exp. 1, the response rates of a greater percentage of *S*s were maintained for a longer period of time by noncontingent presentation of reinforcement than by nonreinforcement. Under extinction, the rate of most *S*s plummeted systematically, which by the fifth or sixth extinction session was rarely more than 15% of the average of the last two preceding VR-15 sessions. In contrast to the lack of behavioral maintenance by extinction is the pattern of responding under the various Acc. FI schedules, which for *S*10, *S*13, and *S*17 displayed no evidence of a decline and for *S*15 and *S*16 stabilized to a value which was about 55 and 30%, respectively, of the average of the last two VR-15 recovery sessions. The animal data of Skinner (1938, pg. 164) and Herrnstein (1968) show that previously conditioned behavior when

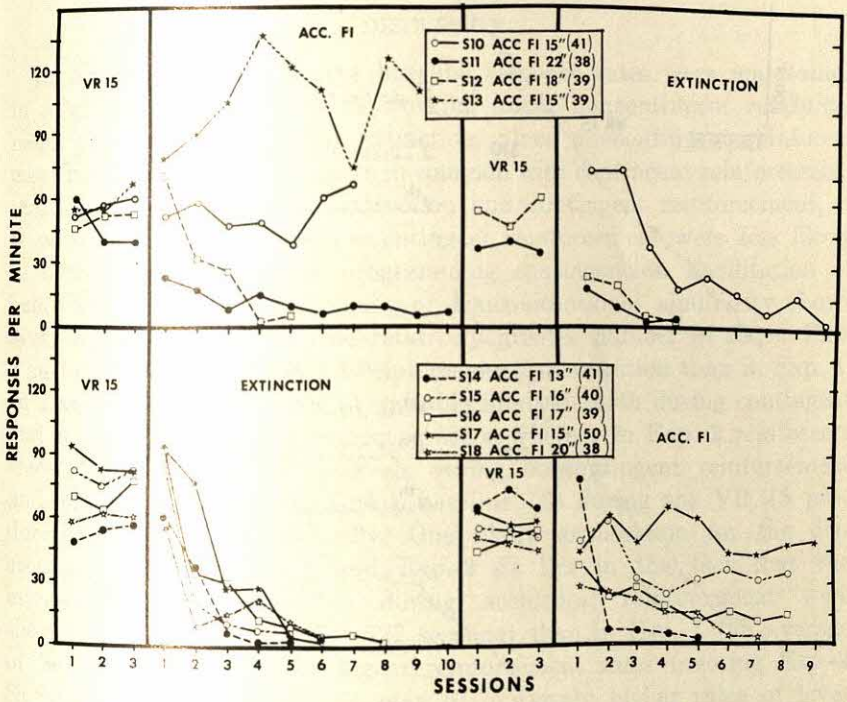


FIG. 3. Changes in response rate over experimental sessions. Only the last three sessions of the original VR-15 baseline and the recovery baseline are shown.

subject to extensive training under Acc. FI schedules ultimately reached an asympotic rate that was about one-third to one-quarter the value obtained with actual contingencies. Needless to say, the frequency and length of exposure to the Acc. FI schedules employed herein is meager in comparison to these animal experiments; had training been extended, the rates of all *S*s might have indeed fallen to similar levels.

Selected cumulative records, reproduced in Figure 4, reveal the within session performance of two of the three *S*s for whom accidental reinforcement had the most durable effects (*S*10 and *S*17); also provided are the records of *S*11 to represent the failure of Acc. FI to maintain previously conditioned VR 15 responding. Two behavioral characteristics of *S*10 and *S*17 during accidental reinforcement is that responding was distributed fairly evenly throughout a session and that shortly before a "free" reinforcement was delivered (pen reset) *S* was usually in the process of lever pressing. *S*13 (records not shown) exhibited a similar pattern except that responding during each FI was more accelerated. On the other hand, *S*s not under accidental control (e.g.,

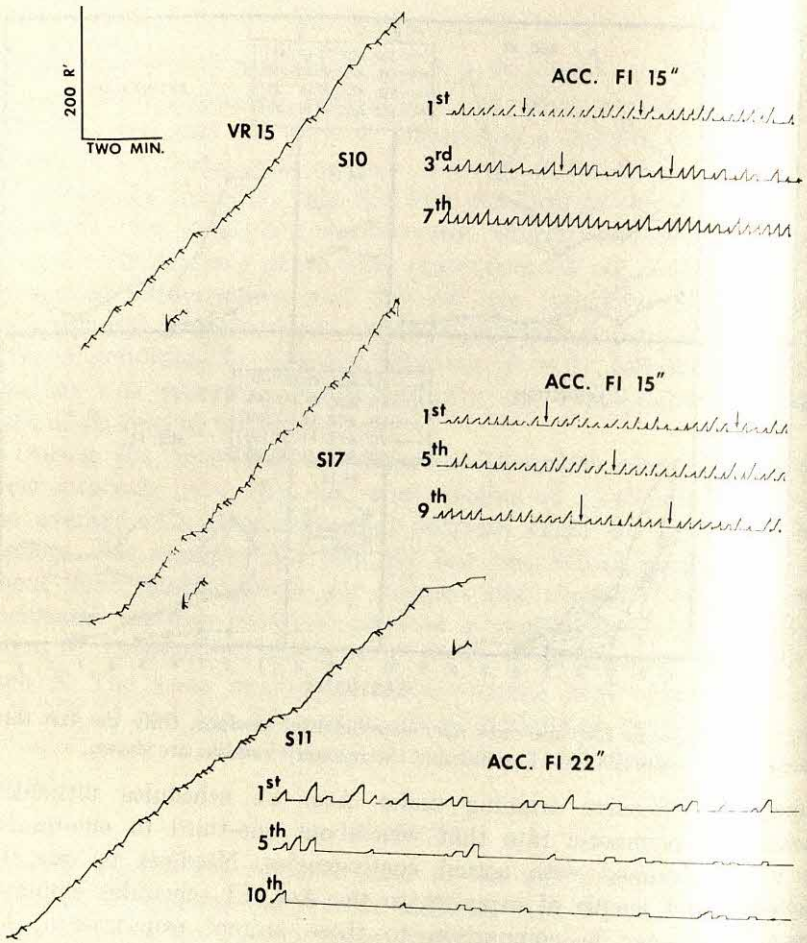


FIG. 4. Cumulative records of the last VR-15 session preceding accidental reinforcement (left side) and selected sessions of the Acc. FI schedule (right side) for three *Ss*. Pen resets at the end of each FI when a reinforcer is delivered. Arrows indicate some, but not all, of the occasions when no responses were made during an interval.

*S*11) responded sporadically throughout a given session and, when responding occurred during an inter-reinforcement-interval, frequently several seconds of nonresponding elapsed before delivery of a reinforcer. Of special interest in Figure 4 is the absence of lever pressing by *S*10 and *S*17 during some intervals (arrows), indicating that there was at least an opportunity for them to attend to changes in the reinforcement contingency, and yet, only some moments after these occasions, these same *Ss* returned to lever pressing.

DISCUSSION

In Exp. 1 and Exp. 2, the baseline response rates were maintained in a greater proportion of *Ss* working under noncontingent reinforcement conditions than under extinction. Since noncontingent reinforcement has more stimulus features in common with contingent reinforcement than is contained between extinction and contingent reinforcement, it is possible that *Ss* during noncontingent reinforcement were less likely to detect a change in the programming contingencies. Facilitation of behavioral maintenance because of trans-situational similiarity, however, cannot account for the relatively greater number of *Ss* in Exp. 2 sustained by noncontingent reinforcement presentation than in Exp. 1. In Exp. 1, reinforcers came at variable intervals both during contingent and noncontingent reinforcement sessions whereas in Exp. 2 reinforcers were delivered at fixed intervals during noncontingent reinforcement and at variable intervals during baseline (*Ss* during the VR 15 produced reinforcers aperiodically. One likely explanation for the discrepancy between Exp.-1 and Exp.-2 *Ss* lies in the fact that the inter-reinforcement intervals during accidental reinforcement were shorter in Exp. 2 (from 13 to 22 seconds) than in Exp. 1 (an average of 30 seconds). Thus, the higher reinforcement rates favoring Exp.-2 *Ss* coupled with these *Ss*' display of uniformly higher rates of lever pressing (because of the previous VR-15 schedule effects) would more strongly safeguard this behavior against extinction or, alternately, make it less likely for Exp.-2 *Ss* to engage in competing behavior at the time of reinforcement.

Associated with reinforcement density is the role of reinforcement delivery and subsequent consummatory activities as discriminative occasions for continued responding. In the response-dependent VR-15 schedule, reinforcement presentations are assumed more likely to function as occasions for lever pressing than in the time-based VI-30 LH-10 schedule; when the schedule conditions are switched, the response-evoking properties of reinforcement delivery should, in the long run, interact more readily with the relatively short inter-reinforcement intervals of the Acc. FI schedule and help reduce the suppressive effects of an otherwise long delay of reinforcement. Although in Exp. 1 and Exp. 2 the absolute rate of responding during baseline predicted adequately *S*'s response level during accidental reinforcement, the stimulus properties of snack presentation may account for some of the inter-subject differences found in both experiments. Spradlin, Girardeau, and Hom (1966) directly tested this stimulus function with retardates in a situation where reinforcement delivery was contingent upon a period

of nonresponding and found wide individual differences in the effectiveness of the reinforcer to repeatedly evoke responding. This function, however, was only moderately correlated with the retardate's response rate engendered by an FR-50 baseline. Worth noting is that when the *S* for whom reinforcement delivery exerted greatest control was later placed on an Acc. FI-30 second schedule for three sessions, a high and sustained output resulted that was similar to that *S*'s baseline performance. At what point in training this *S*, or any of our *S*s who were maintained by accidental reinforcement, learned that reinforcement delivery was a stimulus for lever pressing (if, indeed our *S*s learned this) is presently unclear and deserves further experimental assessment. To check the remote possibility that this function developed *before* baseline training, we presented to three already magazine-trained *S*s (each, about 3 1/2 years old), 35 reinforcements on an accidental FI basis over a range of five to nine sessions (for two *S*s the FI was 15 seconds and for the other *S*, 18 seconds) and recorded the frequency of lever presses during each session. Aside from one session in which 106 presses were tallied, in no other session did any *S* respond more than 20 times.

The programming of response-independent events, arbitrarily designated as an accidental reinforcement schedule, should not be taken as a demonstration of behavioral maintenance because of a capricious pairing of a response with a reinforcer. Although the sustained performances of *S*10 and *S*17 (Figure 4) suggest numerous instances where a lever press could have coincided with the delivery of a reinforcer, a more precise measurement of the time transpiring between the last response and reinforcement is needed as well as the proportion of these responses falling into one of the measured temporal intervals. Should the resulting gradients during noncontingent reinforcement differentiate between *S*s displaying a rapid as against a slow decline in response rate, we would then be able to estimate the significance of accidental pairings toward the maintenance of behavior.

Finally, the fact that noncontingent reinforcement failed to preserve indefinitely the rate of lever pressing of all *S*s indicates the importance of the reinforcement contingency for the maintenance of behavior and supports the bulk of experimental literature on this subject (e.g., Ferster and Perrott, 1968; Hart *et al.*, 1968). On the other hand, the differences between Exp. 1 and Exp. 2 suggest that, at this stage, it is best to couch any remarks about the efficacy of noncontingent reinforcement schedules in terms of the temporal parameters which define the schedule as well as to describe *S*'s experience with and the kinds of performances produced by contingent reinforcement schedules.

REFERENCES

CHURCH, R. M. Systematic effect of random error in the yoked control design. *Psychological Bulletin,* 1964, **62,** 122–131.

FERSTER, C. B., AND PERROTT, M. C. *Behavior principles.* New York: Appleton, 1968.

FLESHER, M., AND HOFFMAN, H. S. A progression for generating variable-interval schedules. *Journal of the Experimental Analysis of Behavior,* 1962, **5,** 529–530.

HART, B., REYNOLDS, N., BAER, D. M., BRAWLEY, E. R., AND HARRIS, F. R. Effect of contingent and non-contingent social reinforcement on the cooperative play of a preschool child. *Journal of Applied Behavior Analysis,* 1968, **1,** 73–76.

HERRNSTEIN, R. J. Superstition: A corollary of the principles of operant conditioning. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application,* New York: Appleton, 1966.

KIMBLE, G. A. *Conditioning and learning.* (2nd ed.) New York: Appleton, 1961.

MILLENSON, J. R. *Principles of behavior analysis.* New York: Macmillan, 1967.

SKINNER, B. R. *The behavior of organisms.* New York: Appleton, 1938.

SKINNER, B. F. "Superstition" in the pigeon. *Journal of Experimental Psychology,* 1948, **38,** 168–172.

SPRADLIN, J. E., GIRDARDEAU, F. L., AND HOM, G. L. Stimulus properties of reinforcement during extinction of a free operant response. *Journal of Experimental Child Psychology,* 1966, **4,** 369–380.

WEISBERG, P., AND FINK, E. Fixed ratio and extinction performance of infants in their second year of life. *Journal of Experimental Analysis of Behavior,* 1966, **9,** 105–109.

WEISBERG, P., AND TRAGAKIS, C. J. Analyses of DRL behavior in young children. *Psychological Reports,* 1967, **21,** 709–715.

WEISBERG, P. Operant procedures for the establishment of stimulus control in two-year-old infants. *Journal of Experimental Child Psychology,* 1969, **7,** 81–95.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 234–249 (1969)

Piaget's Conservation Tasks: The Logical and Developmental Priority of Identity Conservation[1]

FRANK H. HOOPER
West Virginia University

The present research has investigated Elkind's conceptual distinction between identity conservation and equivalence conservation. Although Piaget's description and analysis of conservation acquisition rests primarily on considerations of identity conservation, his assessment format has been exclusively equivalence conservation. The present argument contends that equivalence conservation subsumes identity conservation ability, but also includes a necessary logical deduction sequence for adequate equivalence conservation performance. Accordingly, identity conservation should be an earlier developmental acquisition than equivalence conservation.

Appropriate settings were devised for identity and equivalence conservation assessment which equated the two conservation tasks in terms of perceptual cues and memory requirements. The tasks were administered, in a nonrepeated measures design, to equal numbers of males and females at the kindergarten, first, and second grade levels. A nonparametric multiple contingency analysis revealed that the main effects of conservation task type, age levels, and the sex dichotomy were significant. It was concluded that identity conservation definitely precedes equivalence conservation as a developmental achievement. Accordingly, Piaget's view of identity and equivalence conservation as simultaneous or concommitant developments is erroneous. Identity conservation cannot be adequately assessed in the conventional paired-stimulus format, and care should be taken in considering identity processes or explanatory concepts as a complete description of equivalence conservation achievement.

The Piagetian conservation task format has been the subject of detailed analysis with regard to its logical requirements (Elkind, 1967). Elkind's analysis distinguishes between identity conservation (the conservation of a given stimulus attribute, e.g., substance, weight, or volume, across a reversible transformation and with regard to a single stimulus

[1]This study is based on a dissertation submitted by the author to Wayne State University as partial fulfillment of the requirements for the Doctor of Philosophy degree. The author is grateful to Dr. Irving E. Siegel of the Merrill-Palmer Institute and Drs. Eli Saltz and Charles Solley of Wayne State University for their helpful advice in the preparation of this manuscript. The author also wishes to thank the administrators, teachers, and pupils of the Mt. Clemens, Michigan, Community School District for their courtesy and assistance in the present research.

item alone) and equivalence conservation (the invariance of a quantitative relationship across a reversible transformation of one of the elements of that relationship). This conceptual distinction and the suggested developmental progression from identity to equivalence conservation achievement is the major focus of the present research.

The major aspects of the conventional conservation task may be outlined as follows: Given: Two stimulus items A and B; e.g., containers with equal amounts of small seeds and three separate points or intervals in the conservation setting. *Time 1:* A = B (A and B are judged as containing equal amounts of seeds); *Time 2:* B → C (The contents of B are transferred or transformed to a container of a different shape, C); *Time 3:* A ? C (The *S* is questioned as to the relationship, equality, or difference of amount, between the standard stimulus A and the comparison stimulus C).

If the *S*, when questioned at *Time 3*, responds that A = C, the *E infers* that B and C were in fact judged equal in amount. In contrast, the response A ≠ B results in the *E*'s inference that B ≠ C, hence a nonconservation judgment is assessed.

Note that in the conventional conservation setting outlined above (designated equivalence conservation by Elkind), the *S* is never actually required to judge overtly the relationship of stimuli B and C. Yet, the realization that the property at issue, i.e., the substance-amount, weight or volume of the stimulus array does not alter following the transformation of B to C, is patently essential to a correct solution of the criterial task. Obviously, a *S* who does not judge B = C is not likely to "conserve" the *relationship* of A = B; therefore, A = C as outlined above.

Identity conservation is defined as the realization that the single stimulus transformation B into C does not alter a fundamental property of the quantity in question. Piaget's explanation and description of the processes whereby the child gradually passes from a stage of nonconservation to an intuitive and transitional stage and finally achieves the third stage of completely logically justified conservation performance is based directly on consideration of this identity case. Thus, the three major *post-facto* rationales which are logically adequate and consistent for Piaget, e.g., *addition-subtraction schemas*—"nothing has been added or taken away"; *reversibility*—"if you poured the seeds back to the first container, they would have the same amount"; and *compensation-proportionality* or the compensation of relations—"that glass of seeds is shorter but narrower too," refers to the relationship of B to C as cited above. The later explanation category, the compensation of relations, plays a primary role in Piaget's conceptualization of conservation acquisition. While identity conservation is the focus of Piaget's theoret-

ical explanation and forms the basis for his description of the stages of conservation and quantifying coordination, his assessment format is primarily the paired-stimulus equivalence setting.

Although the *post-facto* explanations or justifications cited above may well be veridical reflections of the processes leading to identity conservation, they do not subsume or describe the total solution to the conventional equivalence conservation problem. In the equivalence case, the child's successful completion of the problem demands two components: (1) A recognition of the B = C invariance, and (2) the logical deduction of A = B, B = C; therefore, A = C, or the maintenance of a quantitative relationship across the irrelevant transformation of one element in this relationship. It is this quantitative relationship which is the essence of the equivalence paradigm, and the only *logical* means to its solution. If the *S* (adult or child) is presented with only A and C, and asked to make a judgment concerning their relative amounts, under- or overestimation of one of the stimuli generally occurs (Elkind, 1967). The *S* is faced with an illusion situation and a logical solution to the judgment questions is not possible without the previous information regarding an equality relationship.

Since equivalence conservation requires the additional deduction sequence, it should be a later cognitive achievement than identity conservation. It also follows that identity conservation should be a necessary but not a sufficient prerequisite condition for adequate equivalence conservation performance. In contrast, Piaget apparently assumes "that identity and equivalence conservation are simultaneous in time, and that the age of equivalence conservation is also the age of identity conservation, so that it is legitimate to infer the age of the latter from the age at which the former is attained" (Elkind, 1967, p. 23). From Piaget's view, there is no contradiction in discussing identity conservation processes while assessing performance in an equivalence setting.

In assessing the relationship of identity and equivalence conservation, certain conceptual and methodological distinctions must be considered. These distinctions center upon the task requirements specific to each conservation setting. Ideally, equivalence conservation should differ from the identity case only with regard to the additional logical deduction sequence. Our problem becomes one of equating the identity and equivalence tasks in terms of all task requirements and perceptual features except for the latter tasks' postulated deduction sequence. Identity conservation involves a single stimulus array, subjected to a number of reversible, physical transformations. For example, colored seeds will be poured from the initial, standard container (A) to a comparison container (C) of taller and narrower dimensions. Following this transforma-

tion, the *S* will be questioned as to the relative amounts of seeds *presently* in container C as compared to the amount *previously* in container A. This conception of identity conservation assessment differs in two important aspects from equivalence conservation as measured in the conventional paired-stimulus format. The latter type of conservation includes the additional deduction sequence described earlier and presents the *S* with an immediate perceptual display in which standard container A and comparison container C "appear" markedly different. This perceptual disparity (heights or levels of seeds in the present setting) is considered an essential aspect of valid conservation assessment for Piaget. The *S* must demonstrate via his logical explanations, that the same amount of material is present in both containers in spite of the perceptual incongruity.

The present identity conservation task requires the *S* to "remember" the situation prior to transformation or pouring of the seeds, e.g., how the seeds appeared in standard container A, level attained, etc. Piaget and Elkind have suggested drawing a reference point or "guideline" on the standard container to assist the *S* in recalling the previous seed level thus equating the identity task with conventional equivalence conservation assessment. Another possible solution is to remove the standard stimulus array from the *S*'s immediate perceptual field; e.g., place the standard container A behind an opaque screen immediately prior to the transformation of the seeds in container B to comparison container C, thus requiring the *S* to "remember" the appearance of container A. This procedure (termed equivalence conservation I) will be used in the present research. It is comparable to the identity conservation task format in terms of memory requirements and perceptual cues presented to the *S*. In addition, a conventional paired-stimulus format (termed equivalence conservation II) as used by Piaget and other researchers will be included. In this case, all the stimulus arrays will be kept in full view of the *S*. Since the two equivalence tasks share the criterial logical deduction sequence, the number of conserving *S*s is not expected to differ for these conditions. Accordingly, in the analysis to follow, if equivalence conservation I is not found to differ significantly from equivalence conservation II, these settings will be combined to form a single, overall equivalence conservation condition.

Since the perceptual incongruity between standard and comparison containers is an integral aspect of the conservation tasks (identity and equivalence cases), it seems appropriate to examine the effects of varying degrees of perceptual disparity upon each of the present conservation task settings. Piaget has cited cases of children at the transitional or intermediate stage II level who conserve when presented with moderate

degrees of perceptual disparity, but who break down and claim different amounts of material are present when extreme transformations are performed. This appears to be one of the major aspects of the transitionary stage description; i.e., "an intermediate stage in which the child achieves conservation in the simple case, but resorts to intuitive perception for the extreme deformations" (Piaget and Inhelder, 1962, p. 21; author's translation). If the amount of conflicting perceptual cues is a major factor, a conservation task with an extreme physical change should be more difficult than one involving a moderate physical change. Accordingly, each of the three major conservation tasks conditions, identity, equivalence I, and equivalence II, will include two degrees of perceptual transformation or deformation, a moderate and an extreme comparison container.

METHOD

Subjects

Subjects were drawn from the kindergarten, first, and second grade classrooms of two elementary schools in Mt. Clemens, Michigan. The schools are in predominantly white, middle-class neighborhoods. Eighteen males and eighteen females from each grade level were randomly assigned to the various conservation task conditions. The age means for the three grades involved were 6, 7, and 8 years, respectively. Teacher-administered Primary Mental Abilities (from 7–1901, revised 1964) test scores indicated IQ means and standard deviations of 111.4 (12.47), 111.8 (12.88), and 109.2 (12.07) for the kindergarten, first, and second grades.

Task Materials

The basic task format included plastic bottles containing equal amounts of small seeds dyed four different colors and glass containers of constant volume and varying shape. The containers used in each of the conservation task conditions are described below.

Container	Dimensions	
	Height	Diameter
Standard 50-ml beaker	2 inches	1¾ inches
Moderate transformation 50-ml graduate cylinder	2 inches	1¼ inches
Extreme transformation 50-ml graduate cylinder	8 inches	¾ inches

In addition, a plastic funnel, an opaque screen, and a tape recorder were used.

Procedure

An attempt was made to create a relaxed atmosphere to insure maximal verbalizations from the children. The *S*s received a preliminary "warm up" experience to familiarize them with the *E*, the containers and seeds, the procedure of seed-pouring, etc. This introductory experience used a 100-ml beaker and two smaller 30-ml containers. After the *S* had poured the white seeds into the 100-ml beaker, the *E* poured unequal amounts of these seeds into the two 30-ml containers. The *S* was then asked, "What can you tell me about the seeds in these two glasses?" Following the *S*'s indication that one container had more seeds than the other, the *E* poured additional seeds into one of the containers, thus equalizing the amounts. The *S* again was asked "What can you tell me about the seeds in these two glasses?" A record was kept of those *S*s who spontaneously used the terms "same" or "equal amounts" of seeds. If the *S* failed to conclude spontaneously that the glasses contained equal amounts of seeds, this was pointed out, with special emphasis on the term "amount." In this manner, at least a rudimentary understanding of the criterial phrases "more seeds" and "same amount of seeds" was insured for each *S*. Following this orientation, the *S* received three experimental trials, each dealing with different colored seeds. In each instance, the *S* personally poured the seeds from the plastic bottles to the standard containers and from these containers to the appropriate comparison glasses. When the initial seed pouring was completed (and an amount-equality judgment agreed upon by the *S* and *E* for the equivalence conservation cases), the transfer to the comparison container took place. Following this, the criterial question specific to each experimental condition was presented, e.g., "Does this glass (gesture to the comparison container) have the *same amount* of seeds or *more seeds* than this glass (gesture toward standard container) has?" The position of the "more" and "same" phrases were alternated from trial to trial in each experimental condition. The *S*'s objective response and the supporting reasons and explanations for this judgment were tape-recorded. All test administration was done in a room separate from the child's classroom setting. Total administration time was 10–15 minutes. Procedures specific to the three major conservation task-formats were as follows:

(A) Identity Conservation—This case centers on a single, fixed amount of seeds which was transferred from the standard 50-ml beaker to either a graduate cylinder 4 inches in height (moderate transformation) or a graduate cylinder 8 inches in height (extreme transformation). Following each transformation, the *S* was asked, "Does this glass (ges-

ture toward comparison container) have the *same amount* of seeds or *more* seeds than this glass (gesture toward the standard container) *had before?"* This procedure was repeated with different colored seeds to yield three separate trials.

(B) Equivalence Conservation I—This case centers upon two identical 50-ml beakers (A and B) with equal amounts of seeds poured into each. Following S-E agreement regarding this equality, container A was moved behind the opaque screen and the contents of container B were poured into container C (representing either a "moderate" or "extreme" transformation as described above). The *S* was then asked, "Does this glass (gesture toward comparison container C) have the *same amount* of seeds or *more* seeds than this glass (gesture toward standard container behind the screen) has?" This procedure was repeated with different colored seeds to yield three separate trials.

(C) Equivalence Conservation II—This case refers to the conventional paired-stimulus conservation setting. Identically shaped containers A and B (50-ml beakers) were filled with equal amounts of seeds. The contents of B were transferred by the *S* to container C (representing either a "moderate" or "extreme" transformation). The *S* was then asked, "Does this glass (gesture toward comparison container C) have the *same amount* of seeds or *more* seeds than this glass (gesture toward standard container A) has?" In this condition, all the containers were kept in full view of the *S*. This procedure was repeated with different colored seeds to yield three separate trials.

Scoring

All the task sessions were tape-recorded in their entirety and the final determination of response categorization was done at a later time. The same scoring criteria were applied to all the experimental conditions, and this included the objective response, more or same amount, and the associated rationales offered by the *S*. In order to pass a given trial, the *S* had to make an equality judgment, e.g., same amount of seeds, and support this estimate with an adequate explanation. The *S*s' explanations were classified as inadequate, e.g., reasons based upon the immediate perceptual features of the task setting or irrelevant statements, or adequate, e.g., *Reversibility:* "You can pour the seeds back into the first glass and they would have the same amount"; *Statement of the operations performed:* "I just poured these seeds into this glass so they still have the same amount"; *Addition-subtraction:* "I didn't add any seeds": or "We didn't spill any seeds" or "That glass (comparison container) didn't have any seeds in it before we poured these seeds into it"; *Compensatory Relations-Proportionality:* "That glass is short and fat,

but this one is tall and skinny so they have the same amount"; *"Sameness" of seeds:* "These are the same seeds here (comparison container) as we had in here (standard container)"; and *References to the previous amount of seeds or the previous state of equality between containers A and B:* "This glass has the same amount of seeds as this one (refers to standard container A and comparison container C) because these two glasses (refers to standard containers A and B) had the same amount of seeds when we started."

Conservation performance in the present study is viewed as an essentially discontinuous, all or none achievement. *Ss* who passed two or three trials are scored as conservers, and this criterion is in accord with the pass/fail specifications of a number of previous investigators, e.g., Kooistra (1964), Shantz and Sigel (1967), Sigel, Roeper, and Hooper (1966). No attempt was made to designate an intermediate category or stage.

RESULTS

Two issues of initial concern are the relationship of equivalence conservation I and equivalence conservation II, and the possibility of distinctive male vs. female performance patterns. Table 1 presents the number of males and females passing the conservation tasks for the various experimental conditions and age-grade levels. Considering, first, the relationship of the equivalence conservation cases Table 1 indicates very little difference in the number of conserving *Ss* for each of these conditions.

For all ages combined, the percentages of *Ss* passing equivalence conservation I and II are noticeably similar, e.g., 44.4% and 41.6%,

TABLE 1

THE NUMBER OF MALES AND FEMALES PASSING THE CONSERVATION TASKS FOR THE VARIOUS CONDITIONS AND AGE LEVELS

		Types of conservation task					
		Identity		Equivalence I		Equivalence II	
Degree of transformation:		Moder.	ext.	Moder.	ext.	Moder.	ext.
Age levels							
Kindergarten	Male	3	2	1	0	0	0
	Female	1	0	1	0	0	0
First Grade	Male	3	2	2	2	2	2
	Female	1	3	0	2	2	1
Second grade	Male	2	3	3	2	3	3
	Female	2	2	1	2	1	1

respectively, (Chi-square = .056, 1 *df*, $p < .90$). Considering the various age levels separately, 66.7% of the second graders passed equivalence I and 66.7% passed equivalence II. For the first graders the values are 50 and 58.3% of the equivalence I and equivalence II tasks (Chi-square = .168, 1 *df*, $p < .70$). Although none of the kindergarten *Ss* passed equivalence II, while 16.7% passed equivalence I, this difference also fails to reach significance (Chi-square = 2.182, 1 *df*, $p < .20$). We may conclude, therefore, that the two equivalence task formats which share the criterial logical deduction sequence do not differ significantly for the present sample. Male/female differences, however, are clearly evident in the present results. Across all conditions and age levels, 64.8% of the male *Ss* and 37% of the female *Ss* were conservers (Chi-square = 7.262, 1 *df*, $p < .01$). Consequently, male and female performances will be considered separately in the following analysis.

For all age levels and task conditions combined, 50.9% of the *Ss* were classed as conservers and 49.1% as nonconservers. A form of analysis appropriate to this pass/fail classification is the multidimensional Chi-square analysis suggested by Sutcliffe (1957) and Winer (1962, pp. 629–632). The present design conforms to Sutcliffe's case 2b with certain restricted classifications and certain parameters estimated from the observed data. The partitioning of the overall Chi-square derived from Table 1 indicated significant main effects for the age-grade levels (Chi-square = 18.301, 2 *df*, $p = <.001$), the type of conservation tasks; identity condition compared to combined equivalence conditions (Chi-square = 5.353, 1 *df*, $p = <.05$), and the male/female dichotomy (Chi-square = 7.262, 1 *df*, $p = <.01$). None of the various interaction values approached significance.

The significant positive relationship between age-grade levels and conservation ability is primarily due to the differential performance of kindergarten *Ss* compared to the first and second graders. For all conservation task conditions combined, 22.2, 61.1, and 69.4% of the kindergarten, first and second grade *Ss* were conservers. Individual comparisons indicated significant differences between the kindergarten and first grade (Chi-square = 11.10, 1 *df*, $p = <.001$) and between the kindergarten and second grade (Chi-square = 16.17, 1 *df*, $p = <.001$). The first grade/second grade differences were not significant, e.g., Chi-square = .551, 1 *df*, $p = <.50$.

Considering the type of conservation task, for all ages combined, 66.7% of the identity *Ss* conserved, compared to 43.1% of those *Ss* in the equivalence conservation conditions. Thus, equivalence conservation is a more difficult behavioral achievement than identity conservations. Although the age/levels conservation task-type interaction values did not

approach significance in the overall contingency analysis, there is some indication the relationship of identity and equivalence conservation alters over the present age range. 9.1, 54.2, and 66.7% of the kindergarten, first, and second grade subsamples passed equivalence conservation. These values may be compared to the percentages of *Ss* passing identity conservation, e.g., 50, 75, and 75% for the kindergarten, first, and second grades, respectively. While all these differences favor the identity conservation case, only the comparison for the kindergarten *Ss* is significant, e.g., Chi-square = 8.026, 1 *df*, $p = <.01$. Thus, the differences between identity and equivalence conservation are most notable at the lower age level, and they tend to diminish as the first and second grade levels are reached. This is especially true of male subsamples in which the numbers of equivalence and identity conservers are markedly different at the kindergarten level, but are very closely matched by the second grade.

As indicated above, male and female performance patterns differ markedly in the present study. Inspection of the pass/fail patterns and the absence of any significant interactions indicates a consistent male performance superiority. Considering the identity/equivalence distinction, male identity conservation performance is uniform (83.3% passing) for all three age-grade levels. Adequate equivalence conservation performance, in contrast, increases sharply from the kindergarten (8.3%) to the first grade (67.7%) to the second grade (91.7%). The female *Ss* show a similar increase in the number of conservers from kindergarten to first grade for both identity and equivalence conservation subtypes, e.g., 20–67.7% for identity and 8.3–41.7% for equivalence conservation. However, the number of female *Ss* achieving either type of conservation does not alter from the first grade to the second grade.

It was expected that greater numbers of conservers would be found for the moderate degree of physical transformation as compared to the extreme transformation cases. This is clearly not the case, for the degree of physical transformation shows a notable lack of influence on both identity and equivalence conservation performance. Across all age-grade levels, task conditions, and subsamples 51.8% of the *Ss* in the moderate transformation conditions conserved as compared to 50% in the extreme transformation cases.

In addition to the preceding considerations, it is possible to examine the three conservation task-types in terms of the logical explanations offered by the *Ss*. The adequate rationales supporting an objective response of "same amount" may be classified into various sub-categories. Table 2 gives the relative percentage of satisfactory explanations for the identity, equivalence I and equivalence II conservation tasks. The per-

TABLE 2
PERCENTAGES OF THE MAJOR ADEQUATE RATIONALE CATEGORIES FOR THE THREE BASIC CONSERVATION TASK CONDITIONS

Explanation category	Type of conservation task		
	Identity	Equiv. I	Equiv. II
1. Reversibility	4.8	4.3	2.2
2. Statement of the operations performed	11.1	—	—
3. Addition-subtraction	50.8	2.2	6.7
4. Compensatory relations-proportionality	25.4	17.4	13.3
5. "Sameness" of seeds	7.9	6.5	—
6. Reference to the previous state	—	69.6	77.8

centages in this table are based upon the total number of primary or final explanations offered by the *S*s in question. Children occasionally offered supplementary explanations in conjunction with the primary justifications represented in Table 2. There were 17, 7, and 9 of these mixed response instances in the identity, equivalence I, and equivalence II conservation cases, respectively.

The patterns for equivalence I and equivalence II are clearly similar. The type of explanation predominantly given in the equivalence conservation formats is a reference to the previous state of equality between stimuli arrays A and B. Compensatory relations-proportionality, reversibility, and addition-subtraction explanations also appear as justification for equivalence conservation but in much lower proportions.

In contrast, the most frequent adequate rationale appearing in the identity conservation case centers on addition-subtraction schemas, e.g., "No seeds were added or taken away." This type of rationale accounts for only 2.2 and 6.7% of the total number of adequate explanations offered for equivalence I and equivalence II conservation. Compensatory relations-proportionality explanations also appear more frequently in identity conservation as compared to the equivalence cases.

DISCUSSION

The present investigation has indicated a conceptual distinction within conventional Piagetian conservation tasks. Identity conservation, as Elkind (1967) has shown, is logically prior to equivalence conservation. The present results indicate that identity conservation is *developmentally* prior to equivalence conservation. Adequate identity conservation performance occurred more frequently than equivalence conservation for all age levels assessed. We may conclude that for discontinuous quantity, identity conservation is an earlier developmental acquisition than equivalence conservation.

The type of explanations offered in support of objective conservation judgments also indicate distinctions between identity and equivalence conservation. As Table 2 indicated, the predominant explanation categories for the identity and equivalence cases were noticeably different. Approximately 50% of the identity explanations focused upon addition-subtraction schemas, e.g., "no seeds were added or taken away." This response has generally been considered an explicit, logically consistent justification and unequivocal evidence of successful conservation performance, e.g., Piaget (1952), Piaget and Inhelder (1962), and Smedslund (1961). Its differential appearance in the present *Ss*' identity explanations adds further support to the developmental priority of identity over equivalence conservation.

The equivalence conservation case, in contrast, is usually "solved" by a reference to the previous state of equality between standard containers A and B. Acknowledging the dangers of an uncritical acceptance of young children's introspective rationales, e.g., Piaget (1954), Werner (1957), it is noteworthy how often the present equivalence subjects offered reasons closely approximating a logical deduction sequence.

For example, Subject D. H. (Male, 1st grade, Equivalence II): "Cause this glass and this glass (gesture to standard containers) *had* the same amount and you poured this glass into there and it's skinnier so it looks like more. No it doesn't have more because these two (gesture toward standard containers) *had the same amount to start out with.*" Subject S. M. (Female, 1st grade, Equivalence II): "Because that one was the same amount and we poured it into here—and the glass is thinner so it makes it look a little bit more but it's still the same amount. Because it *was* the same as that one (gesture to standard) *before* in that glass."

Thus, the primary explanation underlying equivalence conservation focuses upon the original equality relationship which is an integral aspect of the postulated logical argument. As described earlier, it is this logical deduction sequence which formed the basis for a conceptual distinction between the identity and equivalence conservation cases. The present research indicates, in contrast to Piaget's position, that identity and equivalence are not simultaneous or concommitant developments. Piaget is incorrect in attempting to explain equivalence conservation acquisition solely in terms of identity conservation mechanisms or processes. Also, it is clear that the equivalence conservation task format is not the most appropriate means of valid identity conservation assessment.

In view of the heavy emphasis previous investigators have assigned to perceptual cues in conservation task formats. The lack of difference between moderate and extreme transformations is surprising. Although the present results may be specific to the particular container sizes and

materials used, they are in essential accord with Fiegenbaum's (1963) finding that no relationship was present between differential comparison container size and conservation problem solution. If we accept the validity of Piaget's stage descriptions, e.g., Piaget and Inhelder (1962, p. 21), the transitional or intuitive (stage II B) level is not represented by the present *Ss*' performances. Conservation ability was either completely absent or present in the final, generalized stage which is not subject to variations in perceptual disparity or task complexity.

The role of perceptual cues is also involved in the lack of differences found between equivalence conservation I and equivalence conservation II. The screening of the standard stimulus container prior to the transformation had virtually no effect on conservation performance. Apparently, the "image" of the hidden container was sufficient to produce the perceptual incongruity essential to equivalence conservation assessment. There is also the possibility that subjects utilized the empty standard container B as a guideline or reference in judging the A-C relationship. It should be made clear that the present screening procedure differs from Bruner's format which screened the comparison container *and* the results of the transformation, e.g., Bruner, Olver, and Greenfield (1966). Nonetheless, reduction of immediate perceptual cues was not an important factor in the present study.[2]

In addition to these primary considerations, the present research has revealed notably different male and female patterns. While male and female Primary Mental Abilities IQ scores were not significantly different for any age-grade level, male *Ss* were consistently superior for all task conditions and age levels. Not only are the absolute frequency levels different, but, of perhaps greater importance, the relationship between identity and equivalence conservation is not the same for males and females. For the kindergarten girls, identity and equivalence conservation performances are at the same level. The frequency of both conservation types increases from the kindergarten to first grade level with identity showing a marked superiority. Neither case shows any increase from the first to second grade level. Thus, the priority of identity conservation as compared to equivalence conservation is not evident until the first-second grade levels for female *Ss*. For males, in contrast, the

[2] At the conclusion of the equivalence I tasks, each *S* was asked to indicate the approximate level the seeds would reach if the contents of the hidden container A were poured into the empty standard container B. 91.7% of the *Ss* anticipated the level accurately, thus indicating a reasonably clear idea or image of the screened container's level. There was little relationship between correct anticipation and conservation achievement. This agrees with Inhelder's (1965) report that correct anticipations of transformation outcomes precede adequate conservation performance.

greatest difference between identity and equivalence conservation tasks is at the kindergarten level. As mentioned previously, male equivalence conservation performance levels converge with identity conservation at the second grade. In terms of absolute score superiority and developmental progression, the male *Ss* appear more advanced than their female counterparts.

The problem of an adequate understanding and use of the criterion relational terms "more" and "same amount" may also be related to the present sex differences. Thirteen *Ss*, nine girls and four boys spontaneously and appropriately used the term "same amount" in the orientation session which preceded the conservation tasks proper. All four males passed their respective conservation tasks. Eight of the nine females failed to conserve. Apparently, the presence of the relational terms in the *Ss*' vocabulary has a different functional significance for males as compared to females. This lack of prediction between operational usage of the relational terms and conservation performance for the female *Ss* supports an earlier finding by Shantz and Sigel (1967).

It should be emphasized that the present distinction between identity and equivalence conservation differs considerably from Piaget's (1968) recent statements concerning identity concepts. Piaget distinguishes the conservation of *quantitative* invariants, i.e., substance, weight, volume, number, etc., from *qualitative* invariants which are preoperational in nature. The latter category is termed preconservational identity by virtue of its lack of quantitative composition (compensation of relations, for example) and is demonstrated by *Ss* who assert that the amount of water has changed following a reversible transformation but who maintain it is the "same water" as before. Piaget states, "In this case, the invariant is obtained without quantitative composition; there is simply a *dissociation* between a permanent quality (the same water . . .) and the variable qualities (the shape or size), but there is no composition of these variations" (Piaget, 1968, p. 19). This position is essentially similar to the Bruner *et al.* (1966) views on identity concepts as distinct from conservation concepts. The present author's viewpoint in contrast, to Piaget and Bruner, distinguishes between two varieties of conservation acquisition (identity vs. equivalence) *within* the conventional conservation task format.

The present identity/equivalence distinction should be examined in additional content areas and task material settings. Insofar as quantity concepts are concerned, it is probably true that identity conservation performance will vary depending upon the concept area in question and the specific task material used. This has been the case with conventional equivalence conservation, e.g., the horizontal décalage of substance,

weight, and volume, and the greater difficulty presented by continuous stimulus material settings as compared to discontinuous materials. The *relationship* between identity and equivalence conservation, however, would not be expected to vary as a function of the conceptual areas or task materials.

Methodologically, a more stringent test of the logical and developmental priority of identity conservation would involve a repeated measures design with the *Ss* receiving *both* identity and equivalence conservation tasks. Assuming adequate statistical control or evaluation of task carry-over effects a significant number of *Ss* who pass identity and fail equivalence would support the present relationship. Conversely, a sizeable number of *Ss* failing identity and succeeding on equivalence conservation would be a clear-cut refutation of the present position.

A recent assessment of rural Appalachian children (Hooper and Marshall, 1968) evaluated the identity/equivalence conservation relationship utilizing similar task materials and scoring criteria in a repeated measures design. Eighty five- and six-year-old *Ss* were tested, and 25% succeeded on the identity task compared to 13.75% for the equivalence conservation case. More importantly, while 75% of the *Ss* failed both tasks, 13.75% passed both tasks, and 11.25% passed identity and failed equivalence, no *Ss* passed equivalence and failed identity conservation. These results add considerable support to the present contentions.

Finally, the position of identity conservation as a central element within the equivalence task requirements suggests a possible training or enrichment program for promoting conventional, equivalence conservation acquisition. Since identity conservation may be viewed as a necessary but not sufficient prerequisite for equivalence conservation performance, it would seem appropriate to train children on identity task settings. This could focus upon the critical physical alteration of stimulus B to C in the identity setting. A possible training procedure stressing the discrimination of similar spatio-temporal stimulus transformations has been proposed by Watson (1968). General discrimination-memory training of this type has been found to influence quantity and number conservation acquisition, e.g., Shantz and Sigel (1967).

REFERENCES

BRUNER, J. S., OLVER, ROSE R., AND GREENFIELD, PATRICIA M. *Studies in cognitive growth.* New York: Wiley, 1966.

ELKIND, D. Piaget's conservation problems. *Child Development,* 1967, **38,** 15–27.

FEIGENBAUM, K. D. Task complexity and I.Q. as variables in Piaget's problem of conservation. *Child Development,* 1963, **34,** 423–432.

HOOPER, F. H., AND MARSHALL, W. H. *The initial phase of a preschool curriculum*

development project. Final Report, United States Department of Health, Education and Welfare, Office of Education Contract No. OEC-3-7-062902-3070.

INHELDER, B. Operational thought and symbolic imagery. In P. Mussen (Ed.), European research in cognitive development, *Monograph for the Society of Research in Child Development,* 1965, **30,** 4–18.

KOOISTRA, W. H. Developmental trends in the attainment of conservation, transitivity, and relativism in the thinking of children: A replication and extension of Piaget's ontogenetic formulations. Unpublished doctoral dissertation, Wayne State University, 1964.

PIAGET, J. *The child's conception of number.* London: Routledge and Paul, 1952.

PIAGET, J. *The construction of reality in the child.* New York: Basic Books, 1954.

PIAGET, J. *On the development of memory and identity.* Barre: Clark University Press, 1968.

PIAGET, J., AND INHELDER, B. *Le development des quantites physiques chez l'enfant.* (2nd ed.) Paris: Delachaux et Niestle, 1962.

SHANTZ, CAROLYN U., AND SIGEL, I. E. *Logical operations and concepts of conservation in children: A training study.* Final report, project no. 6-8463, United States Department of Health, Education and Welfare, 1967.

SIGEL, I. E., ROEPER, ANNEMARIE, AND HOOPER, F. H. A training procedure for acquisition of Piaget's conservation of quantity: A pilot study and its replication. *British Journal of Educational Psychology,* 1966, **36,** 301–311.

SMEDSLUND, J. The acquisition of conservation of substance and weight in children: II. External reinforcement of conservation of weight and of the operations of addition and subtraction. *Scandinavian Journal of Psychology,* 1961, **3,** 71–84.

SUTCLIFFE, J. P. A general method of analysis of frequency data for multiple classification designs. *Psychological Bulletin,* 1957, **43,** 134–137.

WATSON, J. S. Conservation: An S-R analysis. In I. Sigel and F. Hooper (Eds.), *Logical thinking in children: Research based on Piaget's theory.* New York: Holt, 1968, pp. 447–459.

WERNER, H. The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development.* Minneapolis: University of Minnesota Press, 1957.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 250–255 (1969)

Feedback versus No-Feedback in Testing Children's Knowledge of English Pluralization Rules

BRENDA BRYANT[1] AND MOSHE ANISFELD
Cornell University, Ithaca, New York

Kindergarten pupils were asked to construct plurals for new synthetic singular nouns (Singular-Plural sequence) and to construct singulars for new plurals (Plural-Singular sequence). Half of the *Ss* performed these tasks under a Feedback condition where they were corrected if wrong, and half under a No-Feedback condition where they were not informed whether right or wrong. The *Ss* performed better on the Singular-Plural task than on the Plural-Singular task in the No-Feedback condition. The Feedback condition produced substantial improvement in the latter task and only mild improvement in the former, resulting in equal scores for the two tasks. It was suggested that the Plural-Singular sequence is inherently easier than the Singular-Plural sequence but that the latter is more highly practiced in everyday life. Minimal learning experience (Feedback), however, seems sufficient to equate the two sequences.

This study compares two methods of testing kindergarten children for their knowledge of English pluralization rules. Under one condition of testing, *S* is given feedback information concerning the correctness or incorrectness of his responses, and in the latter case he is supplied with the correct answer. Under the second condition, no such feedback is provided.

The choice of the plural allomorph for regular English nouns is conditioned by the final sound of the singular form of the noun. The /iz/ plural allomorph follows the sounds /s/, /z/, /sh/, /zh/, /ch/, and /j/, as in *dishes* and *watches*. The /s/ allomorph follows the voiceless consonants /p/, /t/, /k/, /f/, and /th/, as in *hats*, *ducks*, and *months*. The /z/ allomorph follows all voiced sounds. The /z/-class can be divided into the following three subcategories: (a) The consonants /b/, /d/, /g/, /v/, and /ð/ (e.g., *hives* and *tithes*) which are constrained on general phonotactic grounds to take a /z/ rather than an /s/. (The sounds listed for the /s/ allomorph are all so constrained.) (b) The consonants /l/, /r/, /m/, and /n/ (e.g., *balls* and *cans*) which if not for the pluralization rule would be free to be followed either by /z/ (e.g., bronze) or by /s/ (e.g., *dance*). (c) Vowels and diphthongs, as in *sofas* and *days*. Category

[1] Now at the University of Minnesota.

(c) is also free evidence the pronunciation of the ⟨s⟩ in the pair *pass-pas* (fathers).

Psychologists have investigated to what extent and in what form young children have internalized these rules (Anisfeld, Barlow, and Frail, 1968; Anisfeld and Gordon, 1968; Anisfeld and Tucker, 1967; Berko, 1958). The present experiment was directly motivated by the Anisfield-Tucker study. They used three production tasks in studying the pluralization rules of 6-year-old children. In these tasks, the children were introduced to synthetic nouns, depicting new creatures, either in singular or in plural form and were requested to produce the complementary form. The children in this study tended to perform better when they were given the plural and had to produce the singular (to be referred to as the Plural-Singular sequence) than when they were given the singular and had to produce the plural (Singular-Plural sequence), although the difference between the two sequences reached significance only in one task. This finding indicates that it was easier for the children to drop off the plural suffix from the inflected form than to add on the plural suffix to the singular. In this study *S*s were given informational feedback after each response: When *S* gave a correct response he was reinforced with "good" and when he gave a wrong response he was told "no," and although a wrong response was recorded in the latter case, *S* was given a chance to try again. When the additional tries did not help *S* to produce the correct response, it was given to him by *E*. A preliminary investigation has shown that the Plural-Singular sequence does not maintain its higher performance in a task which provides no informational feedback.[2] In view of this, it seemed promising to compare directly in one experiment the feedback and no-feedback conditions. The present study was designed for this purpose.

METHOD

Subjects

The *S*s were 72 kindergarten pupils, including 34 boys and 38 girls, enrolled in two schools of the Ithaca (New York) School District. Their mean age was 5 years, 11 months. All *S*s were native speakers of English. Their parents' occupations ranged from unskilled workers to skilled laborers, businessmen, and professionals.

Materials and Procedure

Three sets of 20 synthetic singular nouns of CVC and CV form were constructed, e.g., *tib*, *nup*, and *coe*. One-third of the *S*s was tested with

[2] Mary J. Fratianne participated in this investigation.

each set. All synthetic words constituted permissible sound sequences in English and were pronounceable by 5-year-olds according to Templin (1957). Each of the five allomorph categories described in the Introduction and listed in Table 1 had four representatives—two singular and two plural. For each of the three sets, the 20 words were randomly arranged to constitute one test list. A second list was constructed from the first by changing the plurals to singulars and the singulars to plurals.

There were 20 pairs of cartoon-animal drawings designed to be carriers of the synthetic names. One member of each pair depicted a single animal and the other more than one of the same animal. In the testing procedure, one of the pictures was first shown to *S* and a synthetic name was assigned to it. Following this its counterpart was shown and *S* was requested to produce the name that would be appropriate for it. The standard wording for the Singular-Plural sequence was:

> This is a TIB. [*S* repeated TIB.]
> Several others have come along to join him.
> [The picture with several figures was shown to *S*.]
> Now tell me what you see in the picture.

The expected answer was TIBZ. For the Plural-Singular sequence the questioning proceeded in a similar fashion:

> These animals are called TIBZ. [*S* repeated TIBZ.]
> Some of them have gone away. [The picture of a single figure was shown to *S*.]
> Now tell me what you see in the picture.

The correct answer was TIB.

Before the experiment proper with the synthetic names began, each *S* was acquainted with the task by the use of five real names applied to everyday life objects (e.g., tree, cakes).

The *S*s were tested under one of two conditions: Feedback or No-Feedback. In the Feedback condition, *E* said "good" when *S* replied correctly and "no" when he was wrong. In the latter case, *E* recorded an error, but urged *S* to try again. The *E* would eventually provide *S* with the correct answer. In the No-Feedback condition, *E* said nothing whether *S* had replied correctly or incorrectly and proceeded to the next item. Apart from the presence or absence of feedback, the two conditions were identical in all respects, including specific synthetic words, their order, etc. Testing under the two conditions was alternated between *S*s. They were tested alphabetically with minor rearrangements to pair like sexes under the two conditions. Most of the erroneous responses consisted of *S*'s repetition of the initial form given him by *E*. High reliability of recording of responses was established in previous experiments, and also specifically for the *E* in this experiment (first author).

RESULTS AND DISCUSSION

Table 1 presents the mean number of errors for the two conditions grouped by allomorph category and type of response (singular or plural) required. The data were subjected to an analysis of variance (three factors with repeated measures on two, Winer, 1962, pp. 319–336). The overall mean number of errors in the Feedback condition was significantly smaller than in the No-Feedback condition, $F(1, 70) = 63.6$, $p < .005$. It can be seen in Table 1 that the mean number of errors in the Plural-Singular sequence was much larger in the No-Feedback condition than in the Feedback condition, while that of the Singular-Plural sequence only slightly so and not for all allomorph categories. The Condition-by-Sequence interaction is indeed highly significant, $F(1, 70) = 40.5$, $p < .005$.

The results clearly indicate that the Feedback condition produced a substantial learning effect, especially in the Plural-Singular sequence. To test how rapid this learning was, *Ss*' performance on the first half of the list (10 items) was compared with their performance on the second half (10 items) in each condition for each of the two sequences. This analysis

TABLE 1

MEAN NUMBER OF ERRORS UNDER THE FEEDBACK CONDITION AND THE NO-FEEDBACK CONDITIONS BY ALLOMORPH CATEGORY

No-Feedback		
Allomorph category	Plural–Singular	Singular–Plural
/s/	1.22	.36
/z/(a)—Consonant constrained	1.08	.39
/z/(b)—Consonant free	1.28	.50
/z/(c)—Vowel	1.25	.69
/iz/	1.50	1.36
Overall mean	1.27	.66

Feedback		
Allomorph category	Plural–Singular	Singular–Plural
/s/	.14	.36
/z/(a)—Consonant constrained	.36	.25
/z/(b)—Consonant free	.56	.42
/z/(c)—Vowel	.86	.33
/iz/	.44	.97
Overall mean	.47	.47

Note.—The maximum possible entry in each cell is 2.

shows that of the four cases the only significant improvement from the first half to the second half was in the Feedback condition in the Plural-Singular sequence; mean number of errors for the first half = 1.42, mean second half = .94, $t(36) = 2.66$, $p < .01$. This finding corroborates the results reported in Table 1 which show the strongest effect of Feedback in this sequence.

It may be noted in passing that the Allomorph effect was also significant, $F(4, 280) = 18.4$, $p < .005$. As can be seen in Table 1, the easiest allomorph categories were the two that are phonetically constrained, /s/ and /z/(a), next in difficulty were the phonetically unconstrained categories /z/(b) and /z/(c), and by far the hardest was the /iz/ category. This is generally in accord with previous findings (Anisfeld and Tucker, 1967; Berko, 1958).

The focus of the present study is on the difference between the Feedback and No-Feedback conditions and the interaction of this Condition effect with the Sequence effect. The difference in the effect of Feedback on the two sequences cannot be solely attributed to the fact that the performance on the Singular-Plural sequence was on a high level even without Feedback and it therefore did not leave much room for improvement, since, as can be seen in Table 1, the magnitude of the Feedback effect on the individual allomorph categories is not consistently related to the level of difficulty in the No-Feedback condition. The difference in the effects of Feedback on the two sequences appears to be attributable mainly to the difference in the *nature* of these two tasks. The Plural-Singular sequence calls for deletion of the plural marker whereas the Singular-Plural sequence calls for the suffixation of this marker. Casual observation suggests that new nouns are usually introduced in their singular forms, leaving the speaker the task of constructing the plural forms on his own. The Singular-Plural sequence thus seems more highly practiced in everyday life, and whatever difficulty children still have with it cannot be readily eliminated in the laboratory. The Plural-Singular sequence, on the other hand, presents a more unusual task, according to the above observation, but one which seems inherently easier because the singular form is contained in the plural form. It is likely that for this reason, 20 Feedback trials were sufficient to equate the Plural-Singular sequence with the Singular-Plural sequence. Taking into account the results of the Anisfeld and Tucker (1967) study where they gave *Ss* 36 trials, it seems that with more Feedback trials the Plural-Singular sequence can be rendered even easier than the Singular-Plural sequence. The Feedback procedure thus not only revealed a higher level of performance than the No-Feedback procedure but reflected different aspects of the child's knowledge of English pluralization. The No-Feedback procedure showed

what knowledge the child can immediately demonstrate and the Feedback procedure what knowledge he can readily acquire.

REFERENCES

ANISFELD, M., BARLOW, JUDITH, AND FRAIL, CATHERINE M. Distinctive features in the pluralization rules of English speakers. *Language and Speech*, 1968, **11**, 31–37.

ANISFELD, M., AND GORDON, M. On the psychophonological structure of English inflectional rules. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 973–979.

ANISFELD, M., AND TUCKER, G. R. English pluralization rules of six-year-old children. *Child Development*, 1967, **38**, 1201–1217.

BERKO, JEAN. The child's learning of English morphology. *Word*, 1958, **14**, 150–177.

TEMPLIN, MILDRED C. *Certain language skills in children: their development and interrelationships*. Inst. of Child Welfare Monogr. Series, No. 26. Minneapolis, Minnesota: University of Minnesota Press, 1957.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 256–262 (1969)

The Child's Concept of Proportionality: A Re-examination[1]

KATHERINE J. NELSON,[2] TAMAR ZELNIKER, AND WENDELL E. JEFFREY
University of California, Los Angeles

A partial replication and extension of Bruner and Kenney's (1966) study of the concept of proportionality was run with 5- and 7-year-old children. Experimental *Ss* were given preliminary training to facilitate their understanding of the task and their attention to the relevant cues. Training effectively altered the performance of both the 7- and 5-year-old *E* groups although only the older group was shown to possess the proportionality concept as measured by the criterion task. The results demonstrated the importance of avoiding verbal ambiguity in the investigation of nonverbal cognitive competence.

As an example of the development of the child's ability to handle abstract relationships, Bruner and Kenney (1966) investigated the understanding of physical proportions by young children. They reported that children between the ages of 7 and 11 years pass through several stages of comprehension and do not have an adequate idea of the meaning of a proportional statement until 11 years or older. They proposed that this result was an indication of the young child's inability to base a judgment on the interrelation of two or more stimulus dimensions.

Ginsburg and Rapoport (1967), however, found that both 6- and 11-year-old children were able to compare and to estimate proportions accurately in a nonverbal task. The discrepancy between the results of these two experiments is of interest for both methodological and theoretical reasons. It might be argued that young children have an intuitive grasp of the meaning of proportionality which was tapped in the Ginsburg and Rapoport study, but that they are unable to verbalize it. Alternatively, it might be that the children in the Bruner and Kenney study possessed the concept but were unable to use it due to the particular verbal presentation used. Their *Ss* were asked to tell whether one beaker was "fuller" than another of a different size or shape, con-

[1] This work was supported in part by USPHS Grant MH-6639 to W. E. Jeffrey. We are grateful to Mr. H. Huizing, Principal of Washington School in Redondo Beach for making the children in this study available to

[2] Now at Yale University.

taining a similar or different proportion of water. This instruction may not place sufficient emphasis on the proportionality and therefore may encourage the younger child to make a judgment on the basis of more salient perceptual cues such as relative volume or height. If this were the case then a pretraining situation that emphasized the proportionality implications of the word "fuller" would facilitate performance on the Bruner and Kenney task, and thus give a more accurate estimate of the actual cognitive competence of young children in handling abstract relationships.

The training that was instituted required that the children first appropriately apply the labels "full," "more than half-full," "half-full," and "less than half-full." A short training session followed to assure that the *S*s continued to use these concepts when making comparisons between two beakers.

METHOD

Subjects. Twenty-seven 7-year-old (Mean: 7 years, 9 months; Range: 7 years 2 months to 8 years 1 month) and 18 5-year-old (Mean: 5 years 8 months; Range: 5 years 6 months to 6 years 0 months) Redondo Beach elementary school children were divided into two experimental groups and one control group of 7-year-olds, and one experimental and one control group of 5-year-olds, with nine *S*s per group. The groups were roughly balanced as to sex, *S*s being assigned alternately to groups before being brought to the experimental room. There were 11 girls and 16 boys in the 7-year-old group and nine of each sex in the 5-year-old group.

Materials. A set of five tumblers, each of a different height, circumference and volume, duplicated in all important respects the materials used by Bruner and Kenney. Each tumbler was filled to a predetermined level with Vermiculite (a light sand-like substance).

Procedure. Experimental Groups A (E_A): These groups (one at each age level) were given two sets of training tasks followed by the criterion task. First, a set of six problems requiring *S* to judge the fullness of a single glass was presented twice to each child. The proportions 1/4, 1/2, 3/4 and full were each given one or more times in either of two different sized tumblers presented alternately. Each *S* was instructed as follows: "I am going to show you some glasses filled with sand. I want you to tell me how full the glass is. Just tell me whether it is full, half-full, more than half-full or less than half-full." The *S* was told whether he was right or wrong. Second, each *S* was presented with a set of six problems in which he had to compare two glasses filled to the same proportional level (three problems) or two different proportional levels (three problems). This set was presented twice. The instructions were: "Now I

am going to show you two glasses each time and I want you to tell me how full each one is. How full is this one? (*S* responds). How full is this one? (*S* responds). Is one more full than the other or are they both just as full?" Again the *S* was told whether he was correct or incorrect, and if incorrect he was told why; e.g., "They are both half-full."

The criterion task included seven problems, similar to Bruner and Kenney's Types I, II, and III. Bruner and Kenney's Type IV problems were not used because in their study these were correctly answered by almost 100% of the *S*s. In each problem two different size glasses were used, and the comparisons were as follows: Two half-full glasses (Type I, three problems); two full glasses (Type II, two problems); and one full and one half-full (Type III, two problems); shown in Figure 1. None

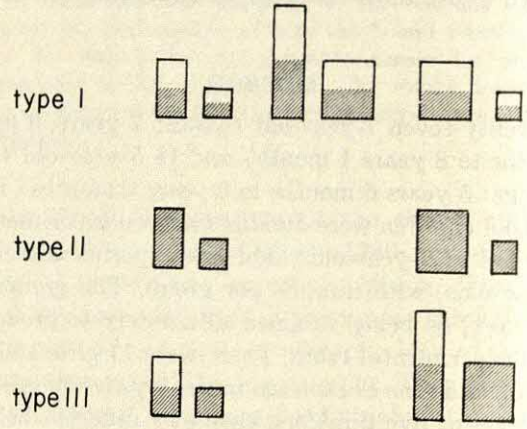


FIG. 1. The pairs of tumblers making up the criterion task as classified by Bruner and Kenney (1966, p. 177).

of the comparisons used in training were duplicated in the Criterion Task. The question asked of *S* was: "Is one glass more full than the other or are they both just as full?" After answering, the *S* was asked to explain why. No feedback was given. Both the answers and explanations were recorded.

Experimental Group B (E_B, 7-year-olds only): In order to be assured that any effect achieved on the criterion task was not attributable simply to the learning conditions provided in the second pretraining task, Group B was given pretraining task 2 without task 1 and *S*s were told whether their judgments were right or wrong, but no explanation was provided for incorrect answers. The Criterion Task followed this training.

Control Groups (C): These groups (one at each age level) received only the Criterion Task. To minimize practice or warm-up effects, they

were given the problems twice but only the scores of the second presentation were used. There was no difference, however, between the first and second presentations for either the answers or explanations at either age level.

RESULTS

Training. The mean number of errors on the task-2 training problems for the experimental groups by age are given in Table 1. Only the 7-year-old Experimental Group A differed significantly from chance performance, and this group also differed significantly from the performance of all other groups on these problems. It should be noted that the mean number of training errors for this group was less than 1.5 on either task, supporting the assumption that training for this group served to clarify the *E*'s instructions rather than to teach a new concept. Thus, for the 7-year-olds, training that demonstrated examples of "full," "half-full," etc. brought *S*s to practically errorless performance immediately. Such training, however, was not successful for the 5-year-old children. Inasmuch as the training in Group B failed to bring the 7-year-olds to an adequate performance level this group was not repeated at the 5-year-old level, as it was deemed unlikely that this limited training would be effective with 5-year-olds and not 7-year-olds.

TABLE 1
MEAN NUMBER OF TRAINING ERRORS FOR EACH *S*

Age	Group	*N*	Training (1)[a] No. problems = 12	Training (2)[b] No. problems = 12
5	E_A	9	5.9	5.3
7	E_A	9	.9	1.4
7	E_B	9		6.7

[a] Difference between 5- and 7-year groups: $t = 5.95$, $p < .001$.
[b] Difference between 5- and 7-year groups: $t = 5.20$, $p < .001$; difference between $7E_A$ and $7E_B$: $t = 7.27$, $p < .001$.

Criterion task. Table 2 shows total mean number of answers and explanations that were correct for both age groups on the Criterion Task. Explanations were considered correct if they referred to the relevant dimension of proportionality (e.g., they are both half-full; one is full but the other is only half-full), rather than to volume or height of sand.[3] All experimental groups differed significantly from their control groups in

[3] Some of the explanations were accompanied by *S*'s pointing to the level of the sand in the glass that had the larger volume of sand in it, but classification was made on the basis of verbal reports only.

TABLE 2
MEAN NUMBER AND PERCENTAGE CORRECT ANSWERS AND EXPLANATIONS FOR EACH *S* ON CRITERION TASK (7 PROBLEMS)

	Group					
	E_A		E_B		C	
Age group	No.	%	No.	%	No.	%
Answers						
5	4.3	61.4			1.8	25.7
7	7.0	100	4.0	57.0	1.7	24.2
Explanations						
5	2.5	35.7			.1	1.4
7	6.0	85.7	1.1	15.7	.5	7.1

number of correct responses ($7E_A$ vs. 7C = $p < .001$ for both answers and explanations; $5E_A$ vs. 5C = $p < .002$ and $p < .001$; $7E_B$ vs. 7C = $p < .02$ for answers and nonsignificant for explanations) and the control groups did not differ from each other. The full-training 7-year-old group (E_A), however, was the only group that could be said to have fully "attained" the concept; the other experimental groups, although they differed significantly from the control groups, performed at a level near chance.

Table 3 shows a comparison of results by type of problem, together with percentage results reported for the same types of problems by Bruner and Kenney for different age levels. The performance for the control groups in this task were similar to those in the Bruner and Kenney experiment, while the performance of the *E* groups was considerably

TABLE 3
PERCENTAGE CORRECT ANSWERS FOR ALL *S*s BY PROBLEM TYPE

Age and group	Type I	Type II	Type III
$5E_A$	29.6	88.8	83.3
5C	11.1	22.2	55.5
$7E_A$	100.0	100.0	100.0
$7E_B$	51.8	50.0	77.7
7C	14.8	11.1	55.5
Bruner and Kenney data:[a]			
(Age group)			
5	12	46	46
7	29	60	50
9	26	44	40
11	69	69	75

[a] In some instances estimated from graphed data.

better. It is noteworthy that Type I problems, which were the most difficult for all groups in the Bruner and Kenney study, were also the most difficult for the children in the present study, but the level of performance on these problems was much higher as the result of training.

DISCUSSION

This experiment demonstrates that here, as in so many other problems in cognitive development, verbal deficiencies or ambiguities in task presentation often mask the actual capacity of the child to perform in the required manner. When 7-year-old children in this experiment correctly understood which dimension in the situation was relevant to the *E*, they performed at a level higher than that previously found for 11-year-olds on a similar task. Their performance on the training task supports the the conclusion that they had the capacity for comparing proportions prior to the experimental situation; training served only to make clear that "fullness" in the *E*'s instructions referred to the proportional aspect of the test materials rather than to other dimensions such as the relative volume or level of the sand in the glasses. Once the instructions were clear there was no difficulty in comparing two proportions and there was no interference from the perceptual cues that otherwise presumably controlled their responses.

Group $5E_A$ and $7E_B$ performed at about chance level on the criterion task but it is interesting to note that the performance of 7C and 5C on the same task was 24 and 26% correct, respectively. These somewhat puzzling results may be clarified by analyzing the children's responses according to problem type rather than across all problems. As can be seen in Table 3, ratio judgments of the control *S*s were not random. Performance of those groups on Type I and Type II problems was between 11 and 22% correct, and about 55% correct on Type III.

As Bruner and Kenney (1966) noted, problems used in this experiment can be ranked according to the number of perceptual cues that mislead the *S*s or support a correct judgment. These cues are the level of sand, the amount of sand (volume), and whether or not the sand reaches the top of the glass. Problem Types I and II both had two misleading cues, but Type II problems were easier than Type I because of the additional "full to the top" cue which was congruent with the correct response. It is therefore reasonable to suggest that the judgment of ratio by the control groups was to a large extent determined by perceptual cues that were basically irrelevant to the concept of proportionality. An analysis of *S*s' explanations gave further support to this conclusion. In each control group all but one of the judgment errors were incorrectly justified in terms of the size of the glasses or the volume of the sand in them.

Five-year-olds with the same training given to the 7-year-olds performed significantly better than control groups but remained at a relatively low level of performance especially on problems of Type I. About 30% of their judgments were not followed by any explanation and for about 30% more incorrect explanation was given. Seventy percent of the errors made by $5E_A$ *Ss* were on Type I problems, and the explanations for those were mainly in terms of misleading perceptual cues. Thus, the results demonstrate that the performance of the 5-year-olds was so strongly dominated by the irrelevant and misleading perceptual aspects of the situation that the brief training used here could not activate the correct ratio concept; or, alternatively, that the correct concept was unavailable to them. However, since only about 50% of the responses made by the 5-year-olds in either training phase were correct, it is possible that their performance on the criterion task might have been improved by more extensive training.

The subjects in Group $7E_B$ performed at about 57% correct on the criterion task, which was significantly better than the performance of 7C on the same task. However, the training they received was not as effective as the verbal training received by $7E_A$ which resulted in near perfect performance in all stages of the experiment.

It can be concluded that when investigating the development of cognitive processes in children it is necessary to avoid semantic ambiguities and to be certain that instructions do not actually mislead *Ss* rather than clarify the task. If the aim is to discover the limits of *Ss'* capability it is important to facilitate *Ss'* attention to the aspects of the problem necessary for correct solution. Children of all ages need to understand the task they are required to perform in order to be able to demonstrate the extent to which they have attained and can use the concept being investigated.

REFERENCES

BRUNER, J. S., AND KENNEY, H. J. On relational concepts. In Bruner, J. S., Olver, R. R., and Greenfield, P. M., (Eds.), *Studies in Cognitive Growth*, 1966. New York: Wiley. Pp. 168–182.

GINSBURG, H., AND RAPOPORT, A. Children's estimates of proportions, *Child Development*, 1967, 38, 205–212.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 263–270 (1969)

Two Attempts to Condition Eyelid Responses in Human Infants

TORU NAITO[1] AND LEWIS P. LIPSITT
Brown University

The conditionability of the eyeblink response using a photocell method with human infants was explored. In one study, an experimental and control group each of 20 *Ss* were studied, the latter group being administered the CS and UCS under a backward conditioning procedure. The results showed a modest conditioning effect at mean age of 40.3 days to tonal stimulation. However, the infants did not produce many CRs, and the CRs did not gradually increase. In a second experiment, differential eyelid conditioning to tonal stimulation was studied in nine babies with mean age of 31 days. The results revealed a differential conditioning effect, although this phenomenon was again unstable. The data indicate the feasibility of studying such conditioning in young infants. Further studies employing similar techniques, and especially involving more trials over a longer age-span, appear warranted.

Eyelid conditionability of infants has been studied by Wenger (1936), Morgan and Morgan (1944), Rendle-Short (1961), and Lintz, Fitzgerald, and Brackbill (1967). However, it is difficult to evaluate some of these studies because of lack of stimulus control, the limited number of *Ss*, and the considerable problem of observing and recording responses reliably. Wenger obtained rather negative results from this particular response. Wickens and Wickens (1940) asserted that the Wenger study, although carefully done, must be criticized because some groups contained only two or three infants and one important control group consisted of but four *Ss*. From their own study, Morgan and Morgan concluded that the normal infant cannot learn this response before 54 days of age, and that all normal children over 65 days of age seem to learn it; in this case, however, the responses were merely observed without instrumentation for recording. Rendle-Short, who used 155 apparently normal children between 5 months of age and 7 years, concluded that children 6 months old and under cannot be conditioned. A positive conditioned eyelid effect was

[1] This study was conducted while the first author, now at Kinjo Gakuin University in Japan, was a Research Associate in the Department of Psychology at Brown University. The research was supported in part by USPHS Grant No. NB-04268 to the second author.

obtained by Lintz *et al.*, who used eight experimental and twelve control infants ranging from 33 to 133 days (median = 69 days) at the start of the study. The literature suggests, then, that eyelid conditioning is typically unstable, if possible at all, in infants under 65 days of age. Recent studies with babies (Lipsitt, 1966) have indicated that increasing evidence of conditioning in the very young infant is sometimes obtained with apparatus and methodological refinements. For this reason, it was felt that further eyelid conditioning studies might clarify the question as to whether the very young can learn this response, and whether their conditioned responses (CR) approach stability.

Techniques employed in eyelid conditioning are rather well developed since there have been numerous methodological studies with adults. However, in considering the use of eyelid conditioning apparatus with infants, further problems had to be resolved, including the comparability of the records with data collected from adults and the bothersome effect of attaching recording implements to the baby. DeLucia's photocell method (1968), developed in the Brown University infant laboratories, does not place undue restrictions upon the *S*'s movements or require that electrodes, paper eyelashes, or any other objects be attached to the eyelid. It was one of the purposes of the present study to test the usefulness of this method with infants.

METHOD

The *S*s were held in an *E*'s arms facing two lamps of 300 watts each placed horizontally 75 cm apart, located about 2 m in front of and 1.7 m above the *S*. The *S* was given a bottle or pacifier if crying or excessive movement occurred during the experiment. The equipment for recording eyeblink has been described completely by DeLucia (1968). The air-puff as the unconditioned stimulus (US) was delivered to the corner of the left eye through a polyethylene tube of 1-mm aperture; its pressure at the source was 5 lb, which reduced to approximately 1 lb at the eye.

A CR was recorded whenever the record showed a deflection of 2 mm or more in the interval 150–500 msec after onset of the conditioned stimulus (CS) during acquisition. A common practice in eyelid conditioning is to record graphic deflections of 1 mm or more from the baseline as CRs, but in babies the response criterion involving deflections of 2 mm or more effectively eliminates small changes of baseline caused by the baby's facial movements.

The *S*s in the first of two studies were 90 infants, 50 of which could not be used either because of procedural difficulties or excessive crying and movement. An experimental group (Group E) and a control group (Group C) of 20 *S*s each, ranging in age from 31 to 55 days of age (mean = 40 days), were studied. For both groups, the CS was a 400

cps tone at 75 db, lasting 1 sec. For Group E, US was presented 500 msec after onset of the CS for 50 msec. For Group C, the CS and US were paired in reverse; US was presented 1500 msec before the onset of the CS (a backward conditioning procedure to determine whether stimulation by tone and at the same time sensitization to the CS might have resulted in increased eyeblinks). Each *S* was given a 40-trial training session followed immediately by 10 extinction trials. During 40 acquisition trials, two CS-alone tests during each ten trials were interspersed. Both groups received intertrial intervals of 5, 10, and 15 sec, averaging 10 sec, arranged according to a fixed schedule during acquisition trials; for test and extinction trials a fixed 10-sec intertrial interval was used.

In a second study, this time of differential conditioning, nine *S*s ranging in age from 29 to 34 days of age (mean = 31) were studied. The CSs were a 3000-cps tone (75 db) and a 300-cps tone (64 db) lasting 550 msec. The US was presented 500 msec after onset of the positive CS for 50 msec, while the negative CS was never reinforced. Intertrial intervals again averaged 10 sec and varied randomly in a prearranged and fixed schedule. All *S*s were given 160 trials with the positive CS and 160 with the negative CS over 4 successive days. Thus 80 trials were given each day. The positive and negative CS presentations were random with the restrictions that no more than three positive CSs nor more than two negative CSs occur in succession. After the first 180 differential conditioning trials, the CS was reversed so that the previously positive CS was no longer reinforced and the previously negative CS was now regularly reinforced. (The reversal took place on the third day, after 20 conditioning trials.) Before reversal, the positive CS was the 3000-cps tone and the negative CS was the 300-cps tone.

RESULTS

Curves of acquisition and extinction of the CR for the two groups of the first experiment are shown in Fig. 1 in terms of CR percentage in each block of ten trials.[2] Throughout the 40 acquisition trials, Group E is consistently above Group C. The former, however, shows no gradual increase, while the latter is flat. Statistical evaluation of the acquisition trials was made by analysis of variance, which revealed a significant overall difference between the groups ($F = 5.05$, $df = 1/55$, $p < .05$). A significant difference for the last 10 acquisition trials (Block 4) was also obtained between the two groups ($F = 4.22$, $df = 1/55$, $p < .05$), although this procedure admittedly capitalizes on chance.

Further analysis was made of the data in Group E by dividing into two subgroups of 10 *S*s (age: 32–38 days) and 10 older *S*s (age: 43–51

[2] The two tests in each block of ten trials produced no reliable effect and are not included in the further analyses.

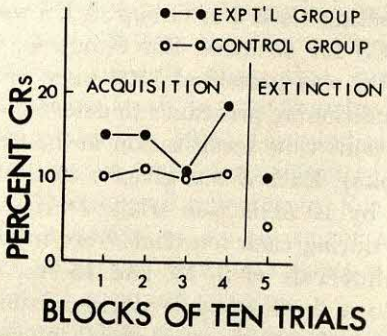


FIG. 1. Curves of acquisition and extinction of the CR for the two groups in terms of percentage of CR in each block of ten trials.

days). During the last 10 trials (Block 4) of 40 conditioning trials, the mean percentage of CRs (27%) by older *Ss* was relatively higher than that of the younger *Ss* (10%), but the difference was not reliable. The reliable difference between Groups E and C could be due in part to the relatively higher level of CRs by the older *Ss* in Group E.

The extinction data showed the same tendency. A significant difference was found between the group means ($t = 2.47$, $df = 38$, $p < .02$). The apparently lower extinction performance than that at the terminal acquisition level was not reliable. Closer analysis of the extinction curves did not indicate any gradual decrease of response. Because the response latency may increase during extinction trials, a CR was recorded whenever the response occurred in the interval 150–1000 msec instead of the interval 150–500 msec; a gradual decrease during the first five extinction trials in Group E was apparent, but no reliable effect was obtained.

The major finding of this classical eyelid conditioning experiment was that of higher response levels during acquisition and extinction trials for Group E as compared with Group C, suggesting that infants might learn eyelid CRs at the mean age of 40 days to tonal stimulation. However, the Group E curve during acquisition showed no gradual increase, and relatively few CRs occurred. It is possible that an insufficient number of trials was administered for producing a stable conditioning phenomenon, or that the infants were not in optimal states for conditioning, or that eyelid conditioning in this age infant is truly unstable.[3]

[3] Another two groups of five *Ss* (which had been used for another experiment before this study) were given 240 trials over 4 days, where daily sessions consisted of 40 acquisition trials, 10 extinction trials, then 10 additional acquisition trials. The mean percentages of CRs per block of 10 trials throughout 50 conditioning trials per day (excluding extinction trials) were 13.6%, 11.2%, 12.4%, and 8% on days 1–4, respectively. Although the conditioning was conducted throughout 200 trials, percentages of CRs did not increase.

The results of the differential conditioning experiments may be quite succinctly summarized. The mean percentage of CRs (10.2%) for the positive CS before reversal was higher than that for the negative CS (7.4%). This difference was not reliable. The mean percentage of CRs (10.2%) for the positive CS after reversal was higher than that for the negative CS (8.6%), again with no difference between the two. However, a significant difference was found between the mean percentage of CRs to the positive CS and the negative CS throughout 160 trials each ($t = 2.55$, $df = 8$, $p < .05$). The result suggests that a differential conditioning phenomenon may have occurred throughout the 160 trials, but differentiation performance stayed low and CRs did not gradually increase (or decrease). It is pertinent to note that the *S*s in the differential conditioning experiment were about 10 days younger than those in the simple classical conditioning experiment. No indications of differential eyelid conditioning seem to have been previously demonstrated with such young babies.

General Comments

Although the conditioning results obtained in the present studies with young infants are modest indeed, an aim of the experiments was to examine the feasibility of studying such processes in infants this young. There are certain special problems in working with infants. One of these is the US intensity and the adaptation of the eyeblink to the air-puff. Ross (1966, p. 37) has called attention to the need for studies of adaptation to the US, both because " . . . it is desirable to use a UCS of sufficient strength to produce a UCR throughout the conditioning session," and because the study of adaptation is of interest in its own right. Figure 2 shows the eyeblink adaptation curves obtained with the infants in Group

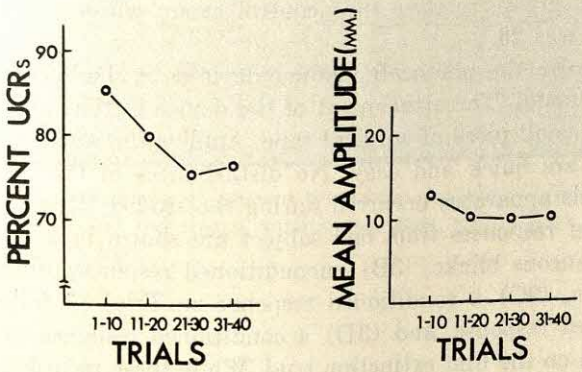


FIG. 2. Frequency and amplitude of URs during training.

C of the first experiment. Throughout 40 trials, the difference between the first 10 trials and the last 10 was a 9% decrease in frequency and a 3.0-mm decrease in amplitude. Neither decline was significant, indicating that rapid adaptation seems not to occur to this particular stimulus.

Further analysis was made of the eyeblink records of the 40 *Ss* in both groups of the first experiment. Excellent readings of the eyeblink records could be obtained between 80 and 85% of the time. The state of the infants during the conditioning procedures was probably not a deterrent to obtaining conditioning.

While the present studies do not provide evidence of robust classical eyelid conditioning in these infants of around 40 days of age, there is nevertheless the suggestion that some degree of such learning was manifested. Coupled with the data of Morgan and Morgan (1944) and more recently that of Lintz *et al.* (1967) that eyelid conditioning can occur in infants by 54 days of age, further parametric and more extensive studies appear warranted. The Lintz *et al.* study shows that CR frequency can, presumably under certain circumstances, be increased markedly in experimental *Ss*; *Ss* in that study at the conclusion of training responded at a level of 90% frequency. The procedure was different from the present study, in that their training was continued to a criterion of 9 CRs out of 10 successive test trials and involved a stronger air puff. Also, test trial presentations of the CS alone were randomly interspersed among conditioning trials in a 1:3 ratio, quite possibly capitalizing upon a partial reinforcement effect. It should be carefully noted, too, that *Ss* in the Lintz *et al.* study ranged in age at the beginning of the experiment from 33 to 133 days, a much wider age span than that of the present study (31–55 days). Using procedures comparable with those of the present classical conditioning study in adults, Naito (in preparation) has obtained an end-of-training level, over a series of 40 trials, of 72% CR frequency in an experimental group, relative to a control group whose comparable percentage CR was 28.

Recording by the photocell method devised by DeLucia (1968) was deemed successful. The attachment of the device to the child's face is effected by a small piece of surgical tape. Application and removal of this attachment are quick and easy. No disturbances of the infants attributable to this apparatus occurred during the studies. Recordings of various types of responses from one subject are shown in Fig. 3, including (3A) spontaneous blinks, (3B) unconditioned responses on the first and second trials, (3C) a conditioned response on Trial 35 followed by the unconditioned response, and (3D) a conditioned response to a CS-alone presentation on the first extinction trial. While these records are specially selected for exemplary purposes, their occurrence supports the credibility

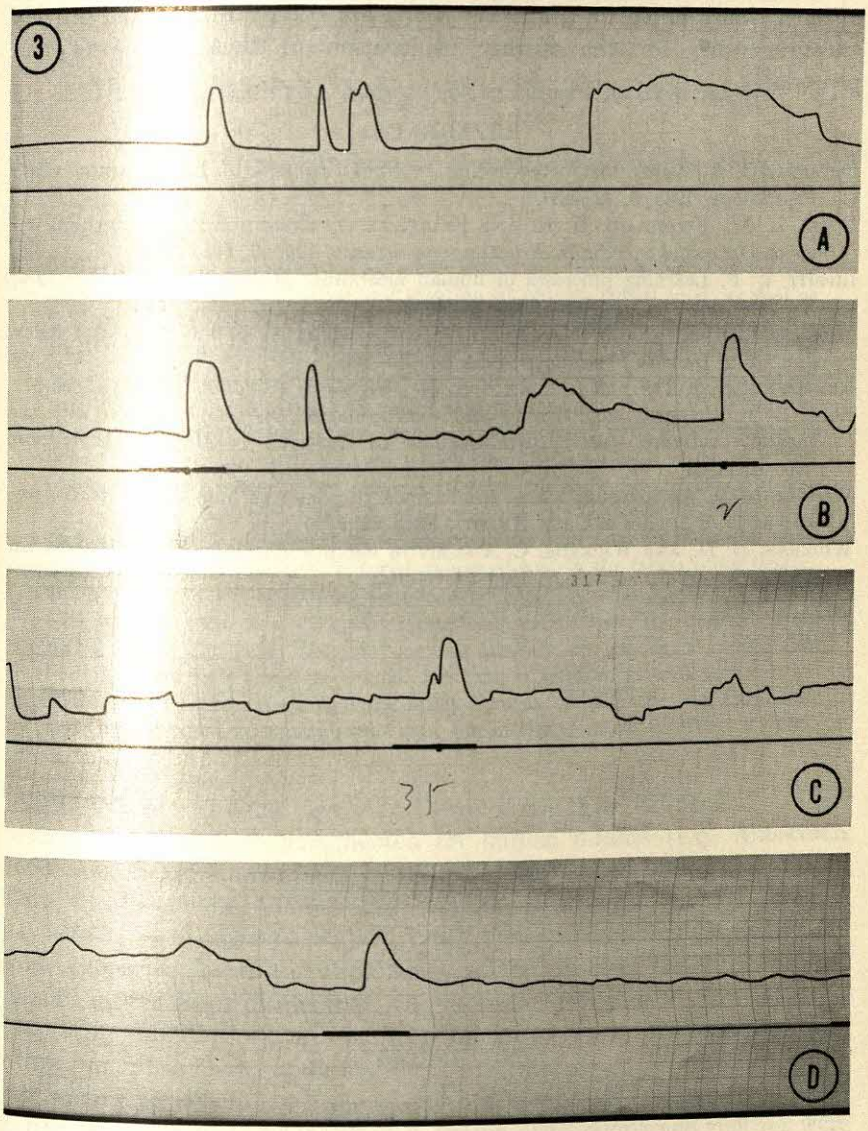


FIG. 3. Exemplary responses of different types from one subject: (3A) Spontaneous blinks; (3B) UR on first and second conditioning trials; (3C) CR on trial 35; (3D) CR on first extinction trial.

of such conditioning in infants as young as 40 days of age and provide encouragement for the further exploration of these processes and procedures.

REFERENCES

DeLucia, C. A. Apparatus for recording eyeblink. *Journal of Experimental Child Psychology*, 1968, **6**, 427–430.

Lintz, L. M., Fitzgerald, H. E., and Brackbill, Y. Conditioning the eyeblink response to sound in infants. *Psychonomic Science*, 1967, **7**, 405–406.

Lipsitt, L. P. Learning processes of human newborns. *Merrill-Palmer Quarterly of Behavior and Development*, 1966, **12**, 45–71.

Morgan, J. J. B., and Morgan, S. S. Infant learning as a developmental index. *Journal of Genetic Psychology*, 1944, **65**, 281–289.

Rendle-Short, J. The puff test. *Archives of Diseases of Childhood*, 1961, 36, 50–57.

Ross, L. E. Classical conditioning and discrimination learning research with the mentally retarded. *International Review of Research in Mental Retardation*, 1966, **1**, 21–54.

Wenger, M. A. An investigation of conditioned responses in human infants. *University of Iowa Studies in Child Welfare*, 1936, **12**, 9–90.

Wickens, D. D., and Wickens, C. A study of conditioning in the neonate. *Journal of Experimental Psychology*, 1940, **26**, 94–102.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 271–283 (1969)

The Infant's Free Entry into a New Environment[1]

HARRIET L. RHEINGOLD AND CAROL O. ECKERMAN[2]
University of North Carolina

The 10-month-old infant's free entry into a new environment was investigated by varying the number and location of objects (toys) in it, as well as the time when they appeared. Each infant was placed with his mother in a small room and allowed access to a larger room (the new environment). In Exp. 1, the effect of a toy in the larger room was compared with that of no toy on the infant's leaving his mother and entering the room. All *Ss* entered the larger room whether it was empty or contained a toy. Those without a toy spent more time with the mother and incidental objects; in contrast, the other *Ss* played with the toy. In Exp. 2, with the same *Ss*, the effect of one toy was compared with that of three toys on the behavior of infants who previously had a toy and those who previously had none. Those who were given toys for the first time entered sooner, stayed in the larger room longer, and played with the toys more than *Ss* who had previously had a toy. Three toys increased the time *Ss* spent away from the mother, the distance they traveled, and the time they played with the toys in the larger room. The findings show that the number and location of objects, as well as the time of their appearance, controlled the infant's behavior in the new environment and call attention to stimuli that lead the infant from his mother.

That infants at some age move away from their mothers to a wider environment has been documented for human infants (e.g., Ainsworth, 1963) and for animal infants in the laboratory (e.g., Bolles and Woods, 1964; Harlow, 1961; Jensen and Bobbitt, 1965) and in the field (e.g., Jay, 1963; van Lawick-Goodall, 1967). The present study investigated some properties of the environment that would lead the 10-month-old human infant from his mother. Of interest also was the effect of these properties upon how quickly he left, how far he went, how long he stayed away, and what he manipulated.

A simple situation composed of mother, infant, two empty rooms, and a few toys provided the laboratory setting. The infant and mother were placed in one of the rooms, and the properties of the other and larger room were manipulated by varying the toys—the time of their appearance and their number and location.

[1] This investigation was supported in part by a Public Health Service Research Career Award (HD-23 620) and a Public Health Service Research Grant (HD-01107).
[2] We thank Don W. Hayne for statistical advice, Richard A. Helwig for technical assistance, and David W. Jacobowitz and Cathy K. Suiter for their services as observers.

In a first experiment, the effect of a toy in the larger room was compared with that of no toy on the infant's leaving his mother and entering the room. The primary question was whether the infant *would* enter the room. In the second experiment, with the same *S*s, the effect of one toy was compared with that of three toys on the behavior of both infants who previously had a toy and those who previously had none. It was expected that the toys would exert a more powerful effect upon the behavior of infants who previously had none. It was also expected that three toys spaced across the room would induce the infants to go farther from their mothers than one toy and to stay away longer.

The present study contrasts with a previous one (Rheingold, 1969) in which infants *placed alone* in the larger room of the present study cried and did not explore, even when the room contained toys. In the present study the infant would start from his mother's side and could if he wished enter the larger room by himself. His entry would be free rather than forced.

The behavior of interest lies within the category of exploratory behavior, both locomotor and investigatory. It also deserves attention as part of the process by which the infant detaches himself from his mother and her near environment.

METHOD

Subjects

Subjects ($N = 24$) were chosen from the population of infants born at the University Memorial Hospital. Selection was on the basis of age and ability to creep, as reported by the mothers. The mothers of those children closest to 10.0 months of age were asked to participate in the study; less than 10% refused.

The age of the subjects (*S*s) ranged from 9.6 to 10.5 months with an average age of 10.0 months. The sample was composed of 14 males and 10 females, all full term; 20 were white, 1 oriental, and 3 Negro. The mean DIQ on the Bayley Infant Scale of Mental Development was 118, with a range from 93 to 140; the mean DMQ on the Bayley Infant Scale of Motor Development was 109, with a range from 88 to 129. The years of education ranged from 10 to 20 years (average, 17) for the fathers and from 8 to 20 years (average, 15) for the mothers. Thus, as a group, the *S*s came from homes that were above average in level of education but not necessarily in income level, since the fathers of 14 *S*s were students.

The records of eight other *S*s, tested in the course of the study, were not used; one *S* did not creep sufficiently well and seven *S*s did not com-

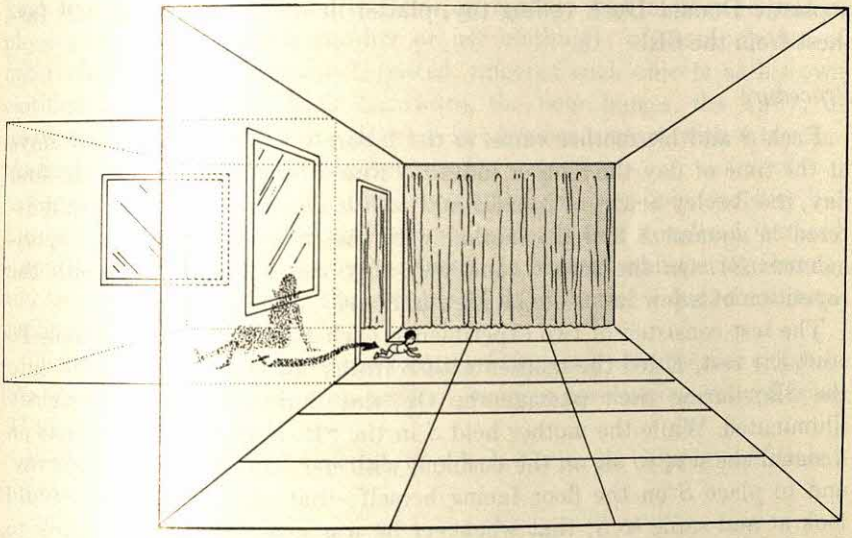


FIG. 1. Floor plan of the experimental space. When present, pull-toy was in cell adjoining SR (column 1, row 2), beads in middle of OF (column 2, row 4), and Donald Duck toy in corner farthest from SR (column 3, row 6).

plete Exp. 2 because of crying. (These *S*s were evenly distributed among experimental conditions.)

Experimental Situation

The experimental situation consisted of two adjoining rooms not seen by the *S*s or their mothers before the experiment (Fig. 1). The larger room, labeled the open field (OF), was approximately 9×18 ft; the smaller one, labeled the start room (SR), was approximately 9×9 ft.

The OF and SR were unfurnished rooms. Their walls, ceilings, and floors were off-white. The floor of the OF was marked off by narrow masking tape into a 6×3 matrix of 18 cells, each approximately 3-ft sq. Off-white draperies covered some walls, concealing windows.

During the experiments the illumination of the OF was greater than that of the SR; incident light on the floor of the OF averaged 22 ft-candles, and on the floor of the SR averaged 4 ft-candles.

Two one-way mirrors gave visual access to the rooms, and a microphone transmitted *S*s' and the mothers' vocalizations to a tape recorder.

For some experimental conditions the OF contained toys, either one or three. For the 1-toy condition, a pull-toy containing colored marbles in a clear plastic ball was placed in the middle of the cell just outside the SR. For the 3-toy condition, the toys were the pull-toy in the same position; a chain of colored wooden beads, placed near the middle of the OF; and

a plastic Donald Duck rolling toy, placed in the corner of the OF farthest from the SR.

Procedure

Each *S* and his mother came to the laboratory on 2 consecutive days at the time of day the mother judged *S* would be most alert. On the first day, the Bayley Scales of Mental and Motor Development were administered to aquaint *S* and the mother with the laboratory and the experimenter (*E*). On the second, and test, day the session began with the repetition of a few items of the Mental Scale.

The test consisted of two experiments, each 10 minutes in duration. To start the test, *E* led the mother with *S* in her arms through the OF into the SR; during their passage the OF was dark and the SR brightly illuminated. While the mother held *S* in the SR, *E* instructed her that on a signal she was to sit on the cushion, with her left side to the doorway, and to place *S* on the floor facing herself; that subsequently she could look at and smile to *S*; that whenever he was near her she could talk to him softly but should limit her speech to short noncommittal remarks; that she could support *S* if he needed help as he stood up near her; and that otherwise she was to allow *S* his freedom, doing nothing to hold him at her side nor to keep him away.

When the mother had placed herself and *S* on the floor, the illumination was decreased in the SR and increased in the OF to the levels reported earlier, and the experiment began. *E* left the SR and joined the observers (*Os*) in the observation room.

At the end of the 10-minute trial *E* instructed the mother over the intercom to close the door to the OF, keeping herself and *S* in the SR. As soon as the toy or toys were positioned in the OF, the mother was directed to open the door, and the second experiment started.

Response Measures

Locomotor activity. On a plan of the floor an *O* traced *S*'s path through the SR and OF. In addition, this *O* and a second *O* independently pressed coded keys activating an event recorder each time *S* crossed the threshold between the OF and SR—that is, when the center of his body crossed the threshold by any mode of locomotion—and each time *S* crossed a line in the OF or touched the lines bounding the perimeter of the room. From the tracing and record of the event recorder several measures were calculated: latency of entering OF, number of entries into OF, duration of entries, duration of time in OF, number of lines crossed, and number and location of different cells entered. No record was made of locomotion in the SR.

Contact and manipulation. The two *Os* independently pressed other

keys to record the onset and duration of contact with mother (*S* was in physical contact with the mother or her clothing); of manipulation of other objects (*S* held, grasped, patted, fingered such objects as his own clothing, the drapes or their hardware, the door hinges, the wall); of manipulation of toys (*S*'s hands were in physical contact with a toy, including the above kinds of manipulation).

Vocal behavior. From the tape recordings, two *O*s independently recorded the number of 10-second periods in which an *S* vocalized (uttered a nonprotest, nondistress sound) or fussed (a broken, whimpering sound), and the latency of fussing. The records of *S*s who cried (loud, continuous wailing) were not used.

The measures reported are the means of two *O*s. Interobserver reliabilities were substantial; Pearson product-moment correlations ranged from .97 to .99 ($p < .001$).

E dictated a running account of *S*'s behavior to supply additional information.

Experimental Design

Experiment 1 tested the effect of a toy in the OF, compared with no toy, on the infant's behavior in leaving his mother. The 24 *S*s were assigned to pairs in order of time of appearance at the laboratory and then allocated by random permutations of two to Group 0 (no toy in OF) or Group 1 (one toy in OF). By inspection, the groups were similar in the subject variables. Statistical testing was by the Wilcoxon matched-pairs signed-ranks test.

Experiment 2 was designed to make two independent tests, first of the effect of previous experience with a toy or no toy in Exp. 1 and second, of the effect of number of toys in the OF (one toy vs. three toys). The *S*s within each group in Exp. 1 were assigned by random permutations of two to either of two conditions, one toy in the OF or three toys in the OF (Table 1). This procedure resulted in the formation of sets of four *S*s, based on order of testing, with each of the four combinations of previous experience and number of toys (0–1, 1–1, 0–3, 1–3) represented within a set.

The effect of previous experience was tested by comparing the behavior of those *S*s with no toy in Exp. 1 (Groups 0–1 and 0–3) with those who previously had one toy (Groups 1–1 and 1–3). Statistical testing was by the Wilcoxon test, comparing within the pairs of *S*s in each set differing in previous experience but not in number of toys (0–1 and 1–1; and 0–3 and 1–3).

The effect of number of toys was tested by comparing Groups 0–1 and 1–1 with Groups 0–3 and 1–3. Statistical tests again were made by the

TABLE 1
DESIGN OF EXPERIMENT 2

Experimental condition	Experience in Exp. 1: Group 0 (No toy)	Experience in Exp. 1: Group 1 (1 toy)	Total
1 Toy	Group 0–1 (6 *Ss*)	Group 1–1 (6 *Ss*)	Groups 0–1 and 1–1
3 Toys	Group 0–3 (6 *Ss*)	Group 1–3 (6 *Ss*)	Groups 0–3 and 1–3
Total	Groups 0–1 and 0–3	Groups 1–1 and 1–3	

Wilcoxon test, making comparisons within the pairs of *Ss* in each set differing in number of toys, but not in previous experience (0–1 and 0–3; and 1–1 and 1–3).

To discover possible interactions between the effects of previous experience and number of toys, the statistic (0–1) − (1–1) − [(0–3) − (1–3)] was calculated for each set of four *Ss* and tested by the sign test. No significant interaction was found on any measure, thus permitting the above comparisons.

All reported p values are for two-tailed tests.

RESULTS

Experiment 1: Effect of a Toy in the Environment

All *Ss* of Groups 0 and 1 left their mothers' sides, crossed the threshold, and without fussing or crying, entered the OF; thus, *Ss* entered the OF whether the toy was present or not. Mean latency of entry into OF was shorter for Group 1 (see Table 2) but not reliably so. Inspection of the data showed that the scores for six *Ss* of Group 0 fell within Group 1's range.

Table 2 presents the other main findings.

During the 10 minutes of the experiment, both groups spent about twice as much time in the SR as in the OF. Although it was expected that the toy would hold Group 1 in the OF longer, no reliable difference was obtained. Eleven of the 12 *Ss*, however, brought the toy into the SR early in the experiment, at a mean latency of 116 seconds.

All *Ss* of both groups returned to the SR at least once and, further, 10 *Ss* of each group then re-entered the OF at least one more time. As Table 2 shows, one *S* in Group 0 entered 13 times and one *S* in Group 1, 11 times!

Group 0 spent more time in contact with the mother and manipulating other objects than Group 1. Group 1 in contrast was manipulating the toy.

TABLE 2
EFFECT OF A TOY IN THE ENVIRONMENT: EXPERIMENT 1

Measure	Mean		Range	
	Group 0	Group 1	Group 0	Group 1
Latency of entry	129	25	14–438	6–61
Time in open field	167	197	10–369	37–369
Number of entries	4.9	3.8	1–13	1–11
Lines crossed	27.1	14.8	2–87	2–44
Different cells entered	6.3	4.3	1–17	1–12
Contact with mother	192	87[a]	7–344	1–374
Manipulation of objects:	100	42[a]	24–190	1–161
In open field	35	19[a]	0–174	0–141
In start room	65	23[a]	11–130	1–84
Latency of contact with toy	—	31	—	9–80
Manipulation of toy:	—	317	—	59–561
In open field	—	105	—	17–282
In start room	—	212	—	0–497

[a] Difference between groups significant at $p < .05$.

Not only did all of them contact it, but as a group they played with it for more than half the trial—one *S* in fact for 561 of the 600 seconds.

In summary, all *S*s entered the larger room, whether or not it contained a toy. The toy evoked prolonged play and reduced contact with the mother and other incidental objects. The toy had no reliable effect, however, upon time spent in the larger room—perhaps because the *S*s carried the toy into the SR.

Experiment 2: Effect of Previous Experience

As in Exp. 1, all *S*s entered the OF. Group 0 entered almost immediately; in contrast, Group 1 took 79 seconds (Table 3).

Now, eight *S*s of Group 0 brought a toy into the SR (at a mean latency of 255 seconds) and six *S*s of Group 1 did the same (at a mean latency of 208 seconds). Despite the similarity in these responses, Group 0 spent more time in the OF than Group 1; in fact, their mean duration was 5 minutes, with one *S* staying in the OF for 578 of the 600 seconds.

Group 0 contacted a toy sooner and played with the toys longer than Group 1. Group 1, in contrast, manipulated other objects more than Group 0.

In summary, Group 0, who now had toys for the first time, entered the OF sooner, stayed in it longer, touched a toy sooner, manipulated the toy or toys longer and other objects less than Group 1. The change did not

TABLE 3
EFFECT OF PREVIOUS EXPERIENCE: EXPERIMENT 2

Measure	Mean		Range	
	Group 0	Group 1	Group 0	Group 1
Latency of entry	8	79[a]	4–19	1–425
Time in open field	302	178[a]	118–578	12–454
Number of entries	2.0	2.7	1–6	1–7
Lines crossed	11.1	14.7	2–39	2–38
Different cells entered	3.7	4.8	1–11	1–10
Contact with mother	103	153	0–351	40–290
Manipulation of objects:	18	104[a]	1–58	1–402
In open field	2	13	0–17	0–57
In start room	16	91[a]	0–58	0–402
Latency of contact with toy	12	122[a]	6–29	7–450
Manipulation of toys:	402	171[a]	7–572	18–451
In open field	254	96[a]	7–554	8–391
In start room	148	75	0–440	0–402

[a] Difference between groups significant at $p < .05$.

reliably affect the duration of time spent with mother or the amount and range of locomotor activity.

It will be noted from Tables 2 and 3 that on several measures Group 0 responded similarly in Exp. 2 to Group 1 in Exp. 1, while Group 1 responded similarly in Exp. 2 to Group 0 in Exp. 1. But since the experimental conditions differed between the two experiments, an analysis of their interaction is not justified.

Experiment 2: Effect of Number of Toys

The *Ss* for whom the OF contained three toys (Groups 0–3 and 1–3) spent more time in the OF, crossed more lines, entered more different cells of the matrix, and played with the toys longer in the OF than the *Ss* (Groups 0–1 and 1–1) having only one toy (Table 4). All *Ss* having one toy made contact with it. Of the 12 *Ss* having three toys, 10 contacted the pull-toy, 10 the beads, and 8 the Donald Duck toy. Seven *Ss* of each group brought a toy into the SR, but for those with one toy the mean latency was 164 seconds and for those with three toys it was 306 seconds.

In summary, three toys spaced across the OF increased the time *Ss* spent in the OF, the distance they traveled, and the time they played with the toys.

Related Findings

Entries and returns. In both experiments, all *Ss* returned to the SR after a period of time in the OF, and subsequently many *Ss* re-entered the OF

TABLE 4
EFFECT OF NUMBER OF TOYS: EXPERIMENT 2

Measure	Mean		Range	
	1 Toy	3 Toys	1 Toy	3 Toys
Latency of entry	40	47	4–344	1–425
Time in open field	184	296[a]	12–463	149–578
Number of entries	2.5	2.2	1–7	1–5
Lines crossed	7.9	17.8[a]	2–26	2–39
Different cells entered	1.9	6.6[a]	1–8	1–11
Contact with mother	160	96	4–351	0–269
Manipulation of objects:	68	53	1–402	1–250
In open field	12	3	0–57	0–23
In start room	56	50	0–402	0–227
Latency of contact with toy	76	58	8–347	6–450
Manipulation of toys:	246	327	7–560	42–572
In open field	118	232[a]	7–409	42–554
In start room	128	95	0–440	0–295

[a] Difference between groups significant at $p < .05$.

several times. Neither the presence of a toy in Exp. 1 (Table 2), nor the addition of toys in Exp. 2 (Table 3), nor the number of toys (Table 4) produced a reliable difference between the groups in the number of entries into the OF.

Not all returns to the SR resulted in contact with the mother. In Exp. 1, of a total of 57 returns for Group 0, only 36 (63%) were followed by contact; of 42 returns for Group 1, 26 (62%) were followed by contact. Thus, on at least a third of the returns, *S*s entered the SR *without* making physical contact with the mother. And, in Exp. 2, of 22 returns by Group 0, only 77% were followed by contact; of 31 returns by Group 1, 71% were followed by contact.

Maintaining visual contact with mother. Infants sometimes entered the seven cells in the OF from which it was judged that they could see no part of their mothers. In Exp. 1, five *S*s of Group 0 and four *S*s of Group 1 entered these cells; in Exp. 2, four *S*s of Group 0 and six *S*s of Group 1 did. Of those who had one toy in Exp. 2, only one *S* entered such a cell; of those who had three toys, nine *S*s entered such cells. These data show that some *S*s traveled to cells from which they could not see the mother whether or not the OF contained a toy. Furthermore, the presence of a toy in the more distant cells appeared to enhance the likelihood that they would break visual contact with the mother.

Many *S*s looked back at their mothers as they entered the OF and then vocalized to her, according to *E*'s dictation. Especially prominent was such behavior in some infants when they had a toy in the environment;

as they grasped the toy they looked at the mother, smiled, and vocalized, as though sharing a pleasant experience.

Vocal behavior of infants. No *S* of either group fussed at the time of entry into the OF. Subsequently, six *Ss* of Group 0 fussed late in Exp. 1 for a short period of time: mean latency was 559 seconds; mean number of 10-second periods was 5 (of the 60 in a trial). In contrast, only two *Ss* of Group 1 fussed; mean latency was 597 seconds; mean number of periods was 2. No fussing occurred while the *Ss* of either group were in the OF.

In Exp. 2, three *Ss* of Group 0 fussed; mean latency was 497 seconds; mean number of periods was 5. In the same experiment, five *Ss* of Group 1 fussed; mean latency was 339 seconds; mean number of periods was 6. In this experiment there was only one *S* who fussed in the OF, a member of Group 1 who had the same single toy again in Exp. 2 (but she was also trying to walk, and fell).

If fussing is considered a sign of mild distress, the situation evoked little distress indeed. What distress there was, occurred late in the experiment, did not last long, and occurred in the OF only once out of 24 possibilities. If *Ss* fussed, then, they fussed in the vicinity of their mothers.

Nondistress vocalizations, in contrast to fussing, were much more common. Each *S* in each group gave such vocalizations during Exp. 1; mean number of 10-second periods in which they occurred was 26 (range, 12–47) for Group 0 and 26 (range, 12–50) for Group 1. All *Ss* also gave some vocalizations in Exp. 2. Mean number of periods was 20 for Group 0 (range, 6–36) and 27 for Group 1 (range, 8–42).

One further analysis remains: a comparison of Exp. 2 data by number of toys in the OF. Of the 12 *Ss* who had one toy, seven *Ss* fussed; of the 12 who had three toys, only one *S* fussed. Nondistress vocalizations were also more frequent among *Ss* with one toy (28) than those with three (19). Toys appeared to suppress fussing as well as vocalizations, a finding previously reported by Rheingold and Samuels (in press).

Mother's verbal behavior. To determine the nature and extent of the mother's verbal behavior, eight tape recordings, randomly selected from each experiment, were analyzed. In 14 of the 16 records the mother talked at some time during the experiment. The total duration of the talking was short (mean = 5.9 seconds, range = 0–21) relative to the trial duration of 600 seconds. The average number of separate utterances during an experiment was 7, with a range from 0–21. Representative of the mothers' speech were such remarks as: "What's the matter," "How are you doing?" and "Careful!" The analysis showed that mothers followed *E*'s directions and that their speech was supportive rather than stimulating.

Effect of S's sex. Sex was not a variable in the selection of *Ss* nor in assignment to experimental groups. Of the 14 males in the sample, seven were assigned by chance to Group 0 and 7 to Group 1; of the 10 females, again by chance, five fell into each group. No reliable differences by sex were found in any of the measures in either experiment. Rather, on some measures the results were surprisingly similar. For example, in Exp. 1 latency to enter OF was 76.2 seconds for males and 77.5 seconds for females; males crossed 20.7 lines and females 21.3 lines; males contacted their mothers on 66% of their returns to the SR and females, on 64%.

DISCUSSION

Ten-month-old infants left their mothers and with no distress entered a new environment whether it was empty or contained a toy. If the environment contained a toy, infants played with the toy and spent less time in contact with their mothers. When toys were added to the environment, in the second experiment, infants who previously had no toys entered faster, stayed there longer, and played with the toys more than infants who previously had a toy. Furthermore, if in Exp. 2 the room contained three toys rather than one toy, the infants traveled farther from their mothers and stayed away longer, playing with the toys in the new environment. In summary, the results showed that the infants entered a new environment by themselves when permitted to do so and when they could leave and return to a familiar social object. Although all the infants entered the new environment, their behavior was controlled by the number and location of stimulating objects, as well as by whether the objects were part of the new environment from the beginning or were added later.

That infants leave their mothers and with no distress go from one room to another is a matter of everyday observation. But that infants crept into the experimental environment of this study and moved freely about with no distress contrasts sharply with the marked distress and almost complete inhibition of locomotion shown by infants *placed* alone in that same environment (Rheingold, 1969). Although the comparison was not made within the same experiment, the methods, observers, source of *Ss*, and experimental environment were closely similar, and the behaviors compared were so consistent within studies that a gross comparison was considered valid. In that study, the results were attributed to the strangeness of the room. The present results suggest that it was not solely the physical properties of the room but the additional conditions of being placed and left alone that provoked the distress and inhibition. In the present study, instead of being placed in the new environment, the infants could enter by themselves; instead of being left alone, they were placed

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 284–295 (1969)

The Organization of Free Recall by Young Children[1]

KATHERINE J. NELSON[2]
University of California, Los Angeles

Subjective and categorical organization of verbal recall by 5- and 8-year-old children was investigated. Fifteen-word lists composed of adult category items, *S*'s own category items, or non-category items were presented for six recall trials following recall on a training list of category labels or non-category words. Improvement over six trials was found for older *Ss* on all lists; the younger groups improved only on the non-category list. There was a parallel increase over trials in measures of organization, both category and subjective. Organization was about the same for both age groups, and type of training had no effect. Changes in recall over trials and with age were attributed to changing strategies of organization rather than to a change in the basic propensity or ability to organize, or to the learning of common modes of organization.

It is frequently assumed that children improve with age in their ability to organize stimulus input; the nature of the improvement, however, has not yet been clearly specified. One possible view is that younger children are passive receivers of information and that, with age, they learn to organize information actively and efficiently. A view more consistent with current developmental theory and empiricial evidence from cognitive and linguistic studies is that the young child actively organizes the environmental input, but that with increasing age he acquires more efficient strategies for doing so. A third possible view is that the apparent improvement of the young child is largely a function of his acquisition of common ways of categorizing information, supplanting more idiosyncratic ones.

The present study investigated these possibilities by examining children's organization of verbal material in free recall. Recent studies of free recall learning with adults have examined the use of conceptual or associ-

[1] Based on a dissertation submitted to UCLA in partial fulfillment of the requirements for the Ph.D. degree. The author thanks W. E. Jeffrey and T. Trabasso for research guidance; Dr. Madeleine C. Hunter, Principal, and the staff of the University Elementary School for providing subjects and research space; and J. Belmont, G. Fein and W. Kessen for helpful comments on a draft of this paper. Computing assistance was provided by the Health Sciences Computing Facility, UCLA, sponsored by NIH Grant FR-3. Preparation of the manuscript was supported by a NICHHD Post-doctoral Fellowship.

[2] Now at Yale University.

ative principles in retrieving related or unrelated words (e.g., Bousfield, 1953; Cofer, 1965; Cohen, 1963; Tulving, 1962). Related words have been found to cluster in recall protocols according to conceptual categories; unrelated words are also clustered in recall, but the basis for the clustering may be evident only to the subject himself. The latter type of clustering has been termed subjective organization (SO) by Tulving (1962). Both types of clustering are assumed to reflect *S*'s reorganization of the input in order to remember it more effectively; improvement in recall is, in fact, found to co-vary with organization.

Studies with children have differed in the extent to which evidence for organization of recall has been found, thus making it difficult to isolate factors which might lead to improvement in such organization. Rossi (1963, 1964; Rossi and Rossi, 1965) found some conceptual organization in children as young as two years. Bousfield, Esterson, and Whitmarsh (1958) reported an increase in category organization with increasing age in school children, and a recent study by Vaughan (1968) found a linear increase in category clustering by *Ss* in grades one to seven. Laurence (1966, 1967), however, has presented negative evidence suggesting a deficiency in both conceptual and subjective organization in young children.

The present study represents an attempt to determine the extent to which young children vary in the type of organization that they impose on the stimulus input according to type of input, and the extent to which such organization could be influenced by training which emphasized categories to be used in the subsequent stimulus lists. To this end, both conceptual and subjective organization were studied within the same lists as well as between groups and at two different age levels. A list of the child's own category items and a list of adult category items were compared with a list of conceptually unrelated words. It was reasoned that if children improve in overall active organizing ability over the age range studied, the younger children should be poorer than the older on all lists and on both organization and recall measures. If children improve in strategy only, older children should show greater improvement on the conceptual lists than on the unrelated list. If their apparent increasing ability were only that of utilizing common organizational frameworks, then the younger children should organize the "child's own" list as well as the older children, although they would show poorer organization scores on the adult category list. A basic assumption was that an age difference in basic memory span capacity would lead to greater total recall scores for older *Ss* on all lists, regardless of organizing ability.

In addition to the age and recall-list factors, two types of training conditions were used. The *Ss* were pretrained on either a list of category labels or a list of unrelated control words. The category words were

labels or code words for the categories used in the subsequent recall lists for each *S* and were thus expected to prime the use of categorical organization if (a) code words or category labels are important to the use of hierarchical or category organization, and (b) the child could use category organization but did not have readily available a code word capable of mediating such organization. The effects of such training would have implications for both a coding theory of memory organization (e.g., Cohen, 1966) and for a mediation deficit hypothesis (e.g., Reese, 1962).

METHOD

Design. The experimental design was a $2 \times 2 \times 3$ factorial with 10 *Ss* per cell and with repeated measures over six recall trials. The three independent variables were Age (5 and 8 years), Training Condition (categorical labels and unrelated words), and Recall Condition (own category, adult category and non-category lists).

Subjects. Sixty *Ss* at each age level from the UCLA Elementary School were assigned randomly to one of the six treatment conditions. The mean age of the younger group was 5 years, 1.4 months, with a range from 4 years, 6 months to 5 years, 7 months; while the mean age of the older group was 8 years, with a range from 7 years, 4 months to 8 years, 9 months.

Stimulus lists. All *Ss* participated in a preliminary session during which their category item responses to ten standard category labels (furniture, clothes, tools, insects, jobs, animals, flowers, vegetables, fruit, colors) were obtained. These categories were selected from those of Cohen, Bousfield, and Whitmarsh (1957) on the basis of their familiarity to young children and their one-word labels. *E* interviewed each *S* individually, asking him to give as many words as he could think of that belonged with each word said by *E*. The *E* then read the one-word category labels in order, giving *S* as much time as he wanted to respond after each stimulus word, and encouraging and prompting as seemed necessary or desirable. All responses were recorded in the order given. (See Nelson, 1968 for summaries of category response data.)

From these responses, a list of category labels for the five categories to which *S* had given the most item responses was selected for each *S*, thus representing his five strongest or largest categories in the set. These five category labels served as the training list for *Ss* in the Category training condition. The *Ss* in the Non-category training condition were given a standard list constructed by matching each of the possible category labels for Thorndike-Lorge frequency (1944, J-count), first letter and number of syllables, and selecting five of these ten at random. Each list was arranged in five different orders such that each word appeared in each serial position only once.

The 15-word recall lists were constructed as follows:

Own-Category (OC). From the five categories to which *S* gave the most responses, his first three item responses for each category were selected for *S*'s recall list. Thus the lists for this group were assumed to represent each child's strongest responses to his strongest categories.

Adult Category (AC). For each of the five categories for which *S* gave the most responses, the three most frequent adult responses from the Bousfield norms (provided that these were not also among the first three responses by *S* in this category) were selected for each *S* in this group.

Non-Category (NC). The three most frequent adult responses from the Bousfield norms for each of the ten categories was matched where possible for Thorndike-Lorge frequency (1944, J–Count), first letter and number of syllables, and a 15-word NC list was then drawn randomly from this pool, with the restriction that no two words belong to the same conceptual category and that none belong to one of the ten categories used for training. The resulting list was: car, nut, block, ring, road, sink, bank, boy, lantern, school, dance, cake, target, cactus, pet.

Each list was arranged in six different orders such that no word followed any other word more than once, and, for the category lists, no word from a given category followed another word from the same category.

Procedure. Each child was tested individually in a room at the school set aside for that purpose.

Immediate memory test. One to five days prior to the recall session, *S* gave his category responses, and at this time was also given an immediate memory test for monosyllabic high-frequency words taken from the Thorndike-Lorge list of 500 most common words, in an attempt to obtain a baseline memory span independent of the experimental manipulations. The test was similar to the forward digit-span of the WISC (1949), ranging from two to seven words. The *S*'s score on this test was the highest number of words repeated correctly in serial order on either the first or second of two possible sets of a given number of words.

Training. Not more than 5 days after the first session, *S* took part in the training and recall session, lasting from 15 to 20 minutes. Training served the dual purpose of providing a warm-up and introduction to the subsequent recall procedure, and of manipulating the category label condition. Training trials proceeded as follows: *E* read the training list at the rate of one word per second and *S* was instructed to recall the words in any order that he could remember them. A different list order was used on each of the first five trials and these were repeated on the second block of five trials if necessary. Training was continued to a criterion of three consecutive correct recalls or for ten trials, whichever occurred first.

Recall. The free recall phase followed training immediately. The *S* was

given the same instructions as for training and was told that now the list would be longer. The list was presented to *S* in a different order on each trial. The *E* read the list at the rate of one word per second and *S*'s recall followed immediately. No time limit for recall was imposed. The *E* recorded responses in writing and by tape recorder; unclear responses were checked against the tape at the end of the session.

RESULTS

Training. Eight-year-olds performed significantly better than five-year-olds in training, in terms of the mean number of words correct on the last three trials (5.00 vs. 4.02), mean number of trials (4.7 vs. 8.9) and number of *S*s who met criterion (60 vs. 16). There were no training score differences between any of the experimental groups within age levels, however. Since, as noted below, training condition had no effect on recall performance at either age level, the age differences in success on training trials are irrelevant to the further results.

Free recall scores. Recall scores were grouped into blocks of two trials to reduce the large trial-to-trial variance of the younger *S*s. A mixed model analysis of variance with repeated measures over three trial blocks and age, training list and recall list as fixed factors, revealed significant main effects of age ($F_{1,108} = 62.51$, $p < .01$) and trial blocks ($F_{2,216} = 57.75$, $p < .01$). Mean scores for each condition are shown in Table 1.

There was a significant interaction of Age × Trial Blocks ($F_{2,216} = 19.52$, $p < .01$) reflecting the fact that eight-year-olds improved more over trials than did five-year-olds. The interaction of Recall List × Trial Blocks was reliable at the .05 level ($F_{4,216} = 2.17$), reflecting the generally greater improvement in recall over trials of the NC lists. This effect was particularly pronounced for the younger *S*s, for whom the only significant improvement over trials came on the NC list ($p < .01$). There

TABLE 1
MEAN NUMBER OF WORDS CORRECTLY RECALLED AS A FUNCTION OF AGE, CONDITION, AND TWO-TRIAL BLOCK

Age	Recall condition	No.	Trial block 1	Trial block 2	Trial block 3
5 Years	OC	20	4.37	4.70	5.08
	AC	20	5.15	5.00	5.08
	NC	20	4.65	6.00	6.35
8 Years	OC	20	6.62	8.60	9.80
	AC	20	6.65	8.30	9.00
	NC	20	5.50	7.80	8.65

were no main effects of recall and training lists. The order of performance level on the recall lists for the older group was in the predicted direction (OC > AC > NC), but the younger groups showed an opposite trend. Thus, total recall was not significantly affected by list composition or by category priming, but increase in recall over trials was affected by the type of recall list.

Category organization. Organization of the AC and OC lists according to the conceptual categories in the list was measured by the ratio of repetition (RR) which is the ratio of number of category clusters to number of words recalled (Bousfield, 1953). Analysis of variance revealed no significant age differences or recall or training list effects with this measure. On the third trial-block, however, a significant ($p < .05$) age difference in RR was found, consistent with the increasing age difference in recall scores on later trials.

Sequential organization. A simple measure of consistency in recall order across trials was devised to serve a function similar to Bousfield's ITR (Bousfield and Bousfield, 1966) and Tulving's SO (Tulving, 1962) measures, since these measures were deemed inappropriate to these data. The SO statistic does not handle intrusions and repetitions in the recall protocols, which occur frequently at this age level and with oral recall. The ITR measure has not been designed to handle blocks of trials larger than two, and young Ss show considerable trial-to-trial variability in both number of words and specific words recalled on adjacent trials. The EITR statistic also varies with length of recall list (Bousfield and Abramczyk, 1966), a disadvantage in making comparisons across ages. Thus a measure that would assess repetitions of word order across all trials and also take into account repetitions of words and word pairs within a trial was needed.

The measure used here, the Repeated Pairs Index (*RPI*), was computed as a ratio of the number of times a given pair of words was repeated over a given number of trials relative to the number of possible times such a pair could be repeated, which was taken to be the number of times that the first word of the pair was repeated. This is computed from a 15 × 15 matrix similar to Bousfield's (Bousfield, Puff and Cowan, 1964) in which all possible first words (A_i) in any pair form one axis and all possible second words (B_j) form the other. The formula for computing *RPI* then is:

$$RPI = \frac{R_{AB}}{\Sigma(A_i - 1)}$$

where $R_{AB} = \Sigma\Sigma\ [(A_i,B_j) - 1]$ for each (A_i,B_j) cell containing at least one entry.

The *RPI* was computed for each *S* for all six trials and for the first and second half of the total trials as an estimate of the increase in this type of organization over trials. Repetitions of words were included and intrusions were ignored. Mean *RPI* scores are given in Table 2.

An analysis of variance on overall *RPI* arcsin transformed scores, with Age, Training List, and Recall List as factors, revealed no significant main effects. The predicted effect that older *Ss* would organize the category but not the non-category lists better than younger *Ss* was supported for the OC list ($p < .01$, one-tailed t test), and the difference was in the same direction for the AC list, thus lending support to the idea that when organization does not depend on particular strategies of organizing, such as the use of common conceptual categories, younger *Ss* organize as well as older *Ss*.

When improvement over trials was analyzed, a significant age difference was observed ($F_{1,108} = 7.04$, $p < .01$), which may be attributed to the greater variability of five-year-olds on the items recalled on adjacent trials that was noted above. The only list on which the five-year-olds improved significantly in *RPI* scores from the first half to the second half of the trials was the NC list, whereas the eight-year-olds showed improvement on the category lists but not on the NC list.

Number and size of categories recalled. One indication of increased organization is the size of categories or groups recalled (Mandler, 1967). An analysis of the number and size of the categories recalled from the OC and AC lists at each age showed that the eight-year-olds consistently recalled more categories on all trials, but that, whereas the five-year-olds recalled as many items per category as the older *Ss* on trial block one, the older *Ss* recalled significantly more items per category on the last trial block ($p < .001$, one-tailed t test). Thus older *Ss* were expanding the size of their categories over trials, while younger *Ss* were not.

TABLE 2
MEAN REPEATED PAIRS INDEX AS A FUNCTION OF AGE AND RECALL CONDITION

Age	Recall condition	No.	(Trials) 1–3	(Trials) 2–4	All trials
5 Years	OC	20	.16	.14	.19
	AC	20	.08	.09	.18
	NC	20	.12	.24	.24
8 Years	OC	20	.20	.23	.26
	AC	20	.11	.17	.20
	NC	20	.18	.17	.23

TABLE 3
CHUNKING RATIOS

Age	Recall condition	No.	Mean immediate memory (IM) score	Mean 1st trial recall score (1st T)	Mean highest recall score (HS)	Chunking ratio	
						1st T/IM	HS/IM
5 Years	C	40	3.95	4.32	7.00	1.14	1.89
	NC	20	3.70	3.90	7.90	1.11	2.23
8 Years	C	40	4.58	5.57	10.60	1.27	2.38
	NC	20	4.95	4.65	9.65	.97	1.99

In an attempt to measure a similar effect in the NC list and to compare the NC and category (C) lists in this respect, Chunking Ratios were devised, so-called in reference to the chunking or unitization theories of Miller (1956) and Mandler (1967, 1968). Mandler has made explicit the relation between size of word group (or chunk) formed in categorizing and remembering verbal materials and increase in hierarchial cognitive organizing processes. The Chunking Ratio (CR) used here was designed therefore to measure the increase in size of word groups over trials using each child's Immediate Memory score (IM) as a baseline indicating number of single items or chunks that can be held in immediate memory at one time. These ratios thus reflect the degree to which words are added (or chunked) above the basic memory span capacity for any *S* on a particular trial. The CRs used here are the ratio between IM score and first trial score and between IM and the highest recall score obtained by *S* on any trial. Table 3 gives data on these ratios.

It should be noted first that on Trial 1 the ratios are very close to 1.00 for all groups, indicating that little improvement over basic memory span took place on the first trial in spite of the additional time available for rehearsing with the longer list. (See Cooper & Pantle, 1967 on the total-time effect on recall increase with adult *S*s.) It can be further seen in Table 3 that, although older *S*s increased their CRs on C lists more than younger *S*s, ($t = 2.42$, $p < .01$), younger *S*s had equal—or higher—ratios on the NC lists. This implies that in terms of increase in memory beyond the immediate memory span younger *S*s were at least equal to older on the NC lists but were inferior on the C lists.

The relation of organization to memory. The RPI and IM scores were not significantly correlated at either age level. However, RPI was positively and significantly ($p < .01$) correlated with the *difference* between the *S*'s highest recall score and IM score at both age levels ($r = .46$ and .31 for five- and eight-year-olds, respectively). This finding supports the

inference from the Chunking Ratios that the relation between the child's organizing ability or strategy and his recall is independent of immediate memory capacity.

DISCUSSION

The most important findings of this research are the differential rates of change over trials of recall by *Ss* at different ages, with different materials, and the consistent relation of organization to increase in recall. It was shown that 8-year-old *Ss* both increased the number of items recalled over six trials and also increased their organization (category and non-category) as shown by organization measures, size of categories and Chunking Ratios. Five-year-olds, however, tended to remain near the level of their first trial performance, at least on the category lists. The fact that the pattern of their performance on the NC list was similar to that for the older group on all lists indicates that there was no age effect on rate of recall improvement or organization *per se*, but that with the materials used in this experiment the younger *Ss* did not employ an effective organizing strategy on the C lists. Thus, two factors, a change in immediate memory capacity and the employment of appropriate organizing strategies, the first an age-dependent variable, the second an age-related but not age-dependent variable, can account for these data. No change in basic organizing ability *per se* need be postulated.

The fact that the eight-year-old *Ss* improved over trials while 5-year-old *Ss* generally did not may be important in understanding the course of cognitive growth. Although this difference can be attributed to changes in organization strategies which enable the older child to amplify his memory span to a greater degree, some of the results may be a function of non-cognitive factors such as motivation or boredom. Whatever the explanation, studies which measure only one-trial performances at this age level will tend to underestimate real differences which may show up only over trials.

The lack of effect of the category label training on any of the dependent variables offers no support for either a mediated association theory or a coding theory which is dependent upon code labels as an explanation of category clustering at this age level. However, further exploration of the code label variable using cued recall experiments might reveal significant effects, such as have been found with adults (Tulving & Pearlstone, 1966; Earhard, 1967).

In this study, IM may be considered an estimate of the child's basic memory capacity, and his organization score may be considered to measure his ability to expand that capacity. Both the increase in recall

and the organization measures reflect the fact that older *Ss* were better able to expand capacity over trials. However, when IM is taken into account, as in the Chunking Ratios, the difference in the two age groups is confined to the category lists only. These data thus confirm the previous suggestion of a difference between category organization and subjective organization as developmental factors (Laurence, 1966; Vaughan, 1968). They suggest, however, that the kind of organization measured by sequential organization scores may be a more primitive or basic type of organization than is category organization, since it is used as well by younger children as by older, while the younger children utilize the inherent category organization very slightly, even when their own category items are used. The fact that the younger children did not show better performance on either recall or organization measures on the list that contained their own idiosyncratic category responses tends to support the conclusion that it is not learning of standard adult categories but the increased ability to use hierarchical organization in this situation that explains the child's increasing recall performance. An alternative interpretation would propose that some unidentified factor was interfering with the performance of the younger children on the category lists.

These findings could support a view that there is a change from a single connection S-R type of cognitive organization to a hierarchical organization after age five (e.g., White, 1965). It should be noted in this connection, however, that all SO measures, including *RPI*, have measured only these single (and uni-directional) connections. If any groups larger than two are being utilized by *Ss*, and if the groups are flexible in direction, SO-type measures will underestimate the organization. Although studies which employ a strategy of allowing *Ss* to group items prior to recall, such as those of Mandler (1967), enable *E* to estimate clustering with non-categorized materials, they are difficult to apply to the ages under study here. Thus further exploration with techniques and measures combining the content-free advantages of the SO-type scores and the multiple-component aspects of category clustering measures are needed to clarify the possibility of a qualitative change in organizing strategy at this age level.

REFERENCES

BOUSFIELD, A. K., AND BOUSFIELD, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, 1966, **19**, 935–942.

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 1953, **49**, 229–240.

BOUSFIELD, W. A., AND ABRAMCYZK, R. R. Sequential ordering in repeated free recall as a function of the length of the stimulus word list. *Canadian Journal of Psychology*, 1966, **20**, 427–434.

BOUSFIELD, W. A., ESTERSON, J., AND WHITMARSH, G. A. A study of developmental

changes in conceptual and perceptual associative clustering. *Journal of Psychology,* 1958, **92,** 95–102.

BOUSFIELD, W. A., PUFF, C. R., AND COWAN, T. M. The development of constancies in sequential organization during repeated free recall. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3,** 489–495.

COFER, C. N. On some factors in the organizational characteristics of free recall. *American Psychologist,* 1965, **20,** 261–272.

COHEN, B. H. An investigation of recoding in free recall. *Journal of Experimental Psychology,* 1963, **65,** 368–376.

COHEN, B. H. Some-or-none characteristics of coding behavior. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5,** 182–187.

COHEN, B. H., BOUSFIELD, W. A., AND WHITMARSH, G. A. Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957, Office of Naval Research, Contract No. Nonr-631.

COOPER, E. H., AND PANTLE, A. J. The total-time hypothesis in verbal learning *Psychological Bulletin,* 1967, **68,** 221–234.

EARHARD, M. Cued recall and free recall as a function of the number of times per cue. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6,** 257–263.

LAURENCE, M. W. Age differences in performance and subjective organization in the free-recall learning of pictorial material. *Canadian Journal of Psychology,* 1966, **20,** 388–399.

LAURENCE, M. W. A developmental look at the usefulness of list categorization as an aid to free recall. *Canadian Journal of Psychology,* 1967, **21,** 153–165.

MANDLER, G. Organization and memory. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation.* New York: Academic Press, 1967. Pp. 328–372.

MANDLER, G. Association and organization: Facts, fancies, and theories. In T. R. Dixon and D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968. Pp. 109–119.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 1956, **63,** 81–97.

NELSON, K. J. Developmental factors in the organization of free recall. Unpublished doctoral dissertation, University of California, Los Angeles, 1968.

REESE, H. W. Verbal meditation as a function of age level. *Psychological Bulletin,* 1962, **59,** 502–509.

ROSSI, E. L. Associative clustering in normal and retarded children. *American Journal of Mental Deficiency,* 1963, **67,** 691–699.

ROSSI, E. L. Development of classificatory behavior. *Child Development,* 1964, **35,** 137–142.

ROSSI, E. L., AND ROSSI, S. I. Concept utilization, serial order and recall in nursery-school children. *Child Development,* 1965, **36,** 771–778.

THORNDIKE, E. L., AND LORGE, I. *The teacher's word book of 30,000 words,* New York: Columbia University Teacher's College, 1944.

TULVING, E. Subjective organization in free recall of "unrelated" words, *Psychological Review,* 1962, **69,** 344–354.

TULVING, E., AND PEARLSTONE, Z. Availability vs. accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5,** 381–391.

VAUGHAN, M. E. Clustering, age, and incidental learning. *Journal of Experimental Child Psychology,* 1968, **6,** 323–334.

Wechsler Intelligence Scale for Children. New York: Psychological Corporation, 1949.

WHITE, S. H. Evidence for a hierarchical arrangement of learning processes. In L. P. Lipsitt and C. C. Spiker (Eds.), *Advances in child development and behavior.* v. 2. New York: Academic Press, 1965. Pp. 187–220.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 296–305 (1969)

The Magnitude of the Orienting Response in Children as a Function of Changes in Color and Contour[1]

CORNELIA DODD AND MICHAEL LEWIS
Educational Testing Service

To investigate the effects of various changes on the magnitude of the OR, two pairs of stimuli were used such that for each pair, each stimulus served as the repeated event for one group of *Ss* and as the altered event for another. It was predicted that stimulus change resulting in increased stimulus intensity would result in a larger OR than stimulus change resulting in decreased intensity. Chromatic and achromatic stimuli were used to test this hypothesis and the results confirmed this notion. The second set of stimuli, varying in contour, tested the hypothesis that changes independent of intensity should affect OR magnitude. The results support this view and suggest that the salience of the change in terms of the organism's hierarchy of interest must be included in a complete analysis of the magnitude of an OR.

The orienting reflex (OR) has been characterized, in part, by response(s) which habituate with repeated presentation of a stimulus and which reappear when the repeated event is altered (Sokolov, 1963). Sokolov further states that response decrement and recovery are mediated by some central process such as memory acquisition or neuronal model formation. When an external event does not match any internal model, central excitation occurs and results in orienting behavior. When the external event matches an internal model, central inhibition occurs and little or no orienting behavior results. Within this theoretical system, rate of response decrement can be viewed as a function of the speed of model acquisition (Lewis, 1968; Lewis & Goldberg, 1969), and amount of response recovery can be seen as a function of the degree of discrepancy between an external event and an internal model (Lewis & Goldberg, 1966).

The theory provides little information for predicting the magnitude of the orienting behavior when the repeated event undergoes various alterations. It is the purpose of this study to explore various stimulus alterations and their effect on the amount of response recovery. It is known, for example, that the OR is affected by changes in terms of intensity

[1]The research was supported in part by Grant Number HD PO 1-HD-01762 and NSF Grant Number GB-8590. Reprint requests should be sent to Dr. Michael Lewis, Center for Psychological Studies, Educational Testing Service, Princeton, N. J. 08540.

(Zimmy & Schwabe, 1966, Bernstein, 1967). Results reported by Lewis, Goldberg, and Rausch (1967) suggest that factors independent of intensity, such as color and content changes, can influence the magnitude of the OR. Thus, it may be the case that such variables as the relevance of the change to the organism's ongoing behavior, and the salience of the change in terms of the organism's hierarchy of interests must be included in any complete analysis of the magnitude of an OR as a result of changes in a repeated event.

In the present experiment, two pairs of stimuli were used such that for each pair, each stimulus served as the repeated event (S_1) for one group of children and as the altered event (S_2) for another. In this way discrepancy was held constant and only the order of stimulus change was varied. It was predicted that stimulus change resulting in increased stimulus intensity or complexity (e.g., additional elements or dimensions) would result in a larger OR than stimulus change resulting in decreased intensity. The first pair of stimuli, changing from chromatic to achromatic as well as in the other direction was a test of the above hypothesis in that the addition or subtraction of color represents, at least, a complexity change. Given no change in stimulus intensity, or complexity, the theory provides no statement as to the effect of the direction of the change on the magnitude of the OR. If there were significant differences in OR magnitude, factors such as stimulus salience would then have to be evoked. The second set of stimuli equal in number of lines but differing in contour was a test of this problem.

METHOD

Subjects

Two groups of children 3½ years of age (±4 months), served as subjects. Group I consisted of 20 *Ss*, 10 boys and 10 girls, and Group II consisted of 32 *Ss*, 19 boys and 13 girls. No sex differences were observed in the data so the data across sex are pooled.

Procedure

The experimental procedures were identical to those described in the Lewis, Goldberg, and Rausch (1967), and Lewis and Goldberg (1969) studies. Briefly, each *S* was seated at a table in a uniform gray room approximately 5 × 5 feet. An adult female sat to the rear and side of *S*. Visual stimuli were presented by rear-screen projection approximately 2½ feet from *S*'s head and at *S*'s eye level.

Figure 1 presents the two pairs of stimuli that were presented. Set A consisted of six presentations of an achromatic (AC) picture of a family

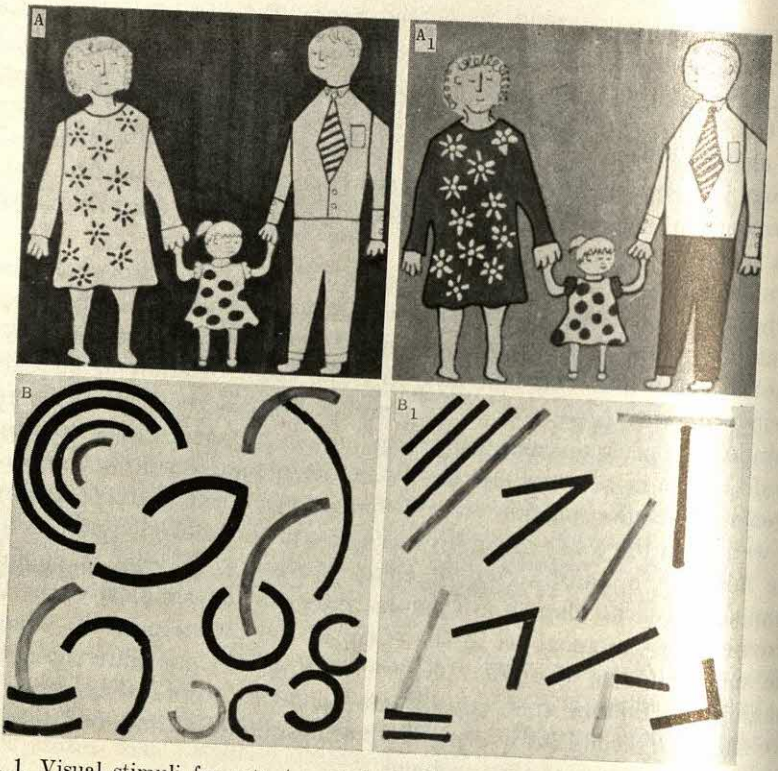


FIG. 1. Visual stimuli for sets A and A_1 (top) and sets B and B_1 (bottom). All stimuli except one (top right) were chromatic.

followed by a seventh presentation of the same picture in color (C); this will be referred to by the notation AC → C. Set A_1 consisted of six presentations of the chromatic picture used in set A, followed by the same achromatic picture: C → AC. Set B consisted of six presentations of a color picture of random curved lines (CL) followed by a seventh trial of a color picture of random straight lines (SL): CL → SL. Set B_1 consisted of the same picture of straight lines as in set B for six trials, followed by the curved-line picture: SL → CL. These stimuli were originally used in the Lewis, Goldberg, and Rausch (1967) study on the effects of various types of stimulus change. Their inclusion in this study was to further explore the effects of color and curvature changes. Since the color dimension has been shown to be an early acquired concept and available at this age (see Lee, 1965) it was thought to see the effects of color and its absence on this type of experimental paradigm. Curvature has also been found to be an effective stimulus change (see Lewis & Goldberg, 1969) and it was thought to further explore this dimension.

Each trial was 30 seconds long with a 30-second intertrial interval. Group I received set A followed by B, while Group II received set A_1 followed by B_1. The interval between the seventh trial of set A (A_1) and the first trial of set B (B_1) varied from 1 to 10 minutes depending on *E*'s judgment of *S*'s restlessness. The amount of rest time was equal across groups. If the child was judged restless he was taken from the room and allowed to play for up to a 10-minute rest period.

Measurement Procedures

As in the earlier work (Lewis, Goldberg, & Rausch, 1967; Lewis & Goldberg, 1969), fixation time was recorded by an observer who was unaware which stimulus was on the screen. An additional observer recorded smiling, pointing, and surprise. The depression or release of a key by the observers activated an event recorder and recorded the occurrence of the behavior. An automatic timer recorded stimulus onset and offset. Interscorer reliability for these measures using the same observers has been consistently high varying from rho values of .68 to .97.

Response decrement was evaluated on the basis of a least square best fit regression curve computed for Trials 1–6 of the form $Y = A + Be^{-Cx}$ (see Lewis & Goldberg, 1969). Response recovery was evaluated on the basis of a comparison of the observed response on Trial 7 with that predicted by the regression equation. In effect, each group served as its own control with behavior on Trial 7 evaluated on the basis of parameters of previous behavior by the same *S*s.

RESULTS

Chromatic and Achromatic Stimuli

Response Decrement

Fixation. The mean total fixation (TF) for each trial for the AC → C and C → AC groups of *S*s and the best fit regression curves for these two groups are shown in Fig. 2A. Both groups showed response decrement fitting the form of a negative exponential function. The best fit curve for the AC → C group was $Y = 7.66 + 23.25e^{-.25x}$ ($F = 67.03$, $p < .005$) and for the C → AC group was $Y = 1.85 + 24.28e^{-.05x}$ ($F = 33.39$, $p < .01$). The raw data indicated more rapid response decrement for the AC → C group. A comparison of the decrement scores (Trials (1–6)/1) for the two groups showed a significant difference (Mann-Whitney *U* test, $Z = 2.00$, $p < .05$), with the AC → C group showing the greater decrement. Moreover, the mean TF over Trials 1–6 was significantly different for the two groups (*U* test, $Z = 3.19$, $p < .001$).

Smiling. Figure 2B presents the mean smiling time for each trial and

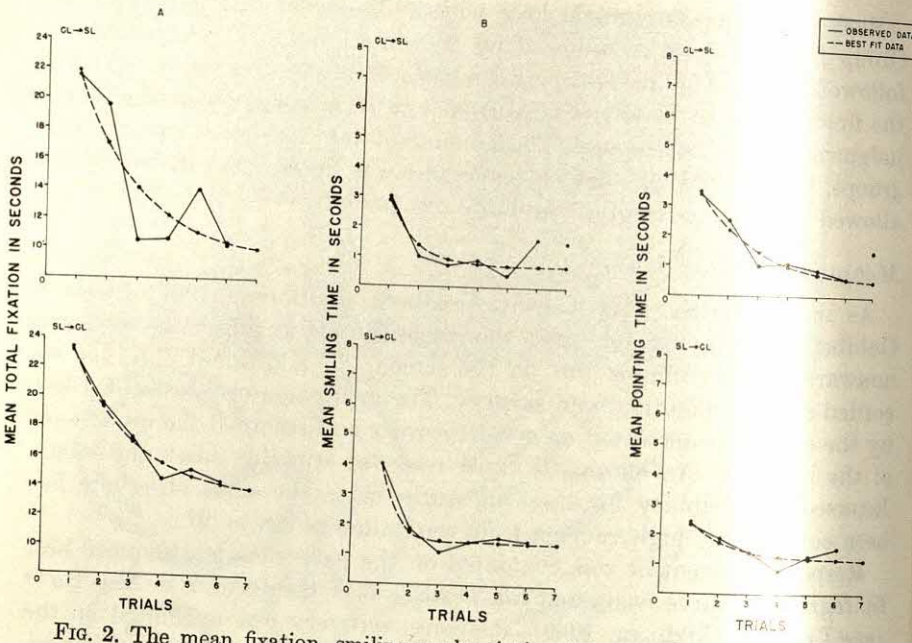


FIG. 2. The mean fixation, smiling, and pointing data and the best fit regression curves for the AC→C and C→AC groups.

the best fit curves for the two groups. Like fixation, the smiling data for both groups showed response decrement fitting a negative exponential function. The best fit curves were $Y = 1.36 + 6.60e^{-.5x}$ ($F = 14.64$, $p < .05$) for the AC→C group, and $Y = 1.10 + 7.73e^{-.75x}$ ($F = 80.64$, $p < .001$) for the C→AC group. Unlike fixation, the response decrement scores for smiling were not significantly different for the two groups (U test, $Z = 0.08$).

Pointing. The pointing data were less consistent than the fixation and smiling data (see Fig. 2C). This was attributable to the fact that fewer *S*s contributed to the data; there were a number of *S*s who either did not point at all, or did not point for a majority of the trials. Consequently, these data did not fit either linear or curvilinear regression curves, and did not show response decrement.

Response Recovery

Fixation. Both the AC→C and C→AC groups showed response recovery on Trial 7, although the recovery was much greater for the AC→C group (see Fig. 2A). The observed mean TF on Trial 7 was significantly different from the points predicted by the regression curves for both groups. For the AC→C group, the observed TF represented a distance of 9.77 standard deviations from the predicted point ($p < 1 \times$

10^{-22}),[2] and for the C → AC group, 4.06 standard deviations from the predicted point ($p < 2 \times 10^{-5}$). In order to compare the differential recovery between the two groups, it was necessary to use difference scores. This difference score was computed by subtracting the mean of the last 3 trials from the value on Trial 7. An alternative procedure would be to obtain the 7- minus 6-Trial differences. While both procedures generate the same results, the former utilizes more trials and reduces chance trial variations. Using the 7 minus the mean of the last three trials resulted in a significant recovery difference between the groups with the AC → C showing greater recovery (U test, $Z = 3.68$, $p < .001$).

Since the TF on Trial 7 approached the maximum TF possible on any one trial (30 seconds), the data were analyzed further to determine if a ceiling effect had limited the fixation on the violation trial. The latency to the first fixation for any trial sets a functional limit on the total amount of fixation for that trial; e.g., if the latency is 10 seconds, the maximum TF possible in a 30-second trial is 20 seconds. The mean latencies were 1.5 seconds for the AC → C group and 2.4 seconds for the C → AC group and the difference between them was not significant. Since the mean TF on Trial 7 for the C → AC group was 24.3 seconds and therefore well within the functional limit for that trial, it appears that the lower recovery for this group is difficult to attribute to a ceiling effect, but rather appears dependent upon the nature of the stimulus and the experimental procedure. Alternative hypotheses about C → AC and AC → C difference are presented in the discussion section.

Smiling. The mean smiling time on Trial 7 also showed response recovery (see Fig. 2B). The observed mean smiling time on Trial 7 represented a distance of 7.75 standard deviations from the predicted point for the AC → C group ($p < 5 \times 10^{-15}$), and 8.08 standard deviations ($p < 5 \times 10^{-16}$) for the C → AC group. A comparison of the difference scores data showed that the AC → C group recovered more than the C → AC group (U test, $Z = 2.04$, $p < .05$).

Pointing. Since the pointing data did not fit regression curves, only change scores were possible to observe. A comparison of the differences between the mean of the last three trials and Trial 7 indicated that both groups showed recovery on Trial 7 (AC → C group: $Z = 1.45$; C → AC group: $Z = 2.41$, $p < .01$). A comparison failed to indicate any significant difference between the two groups.

Surprise. Not enough surprise was recorded to be analyzed statistically. In both groups, all recorded instances of surprise occurred on the violation trial, three in the AC → C group, and six in the C → AC group.

[2] Because response recovery was predicted, one-tail tests were used. For all other tests of significance, two-tail probability statements were made.

Straight and Curved Line Stimuli

Response Decrement

Fixation. The mean TF by trial for the CL → SL and SL → CL groups and the best fit regression curves are shown in Fig. 3A. Both groups showed response decrement fitting the form of a negative exponential function. The best fit curve for the CL → SL group was $Y = 9.41 + 20.51e^{-.5x}$ ($F = 12.40$, $p < .03$), and for the SL → CL group was $Y = 13.40 + 16.24e^{-.5x}$ ($F = 86.67$, $p < .001$). There was no significant difference in the decrement scores (Trials (1–6)/1) for the two groups, nor in the mean TF over Trials 1–6.

Smiling. Both groups showed decrement in the smiling response of the negative exponential form (see Fig. 3B). The best fit curves were $Y = 0.76 + 7.35e^{-1.25x}$ ($F = 13.32$, $p < .03$) for the CL → SL group, and $Y = 1.43 + 21.54e^{-2.1x}$ ($F = 97.49$, $p < .001$) for the SL → CL group. The decrement scores for the two groups were not significantly different.

Pointing. The pointing data are shown in Fig. 3C and unlike the data for sets A and A_1, the data significantly fit negative exponential regression curves. For the CL → SL group, the best fit curve was $Y = 0.34 +$

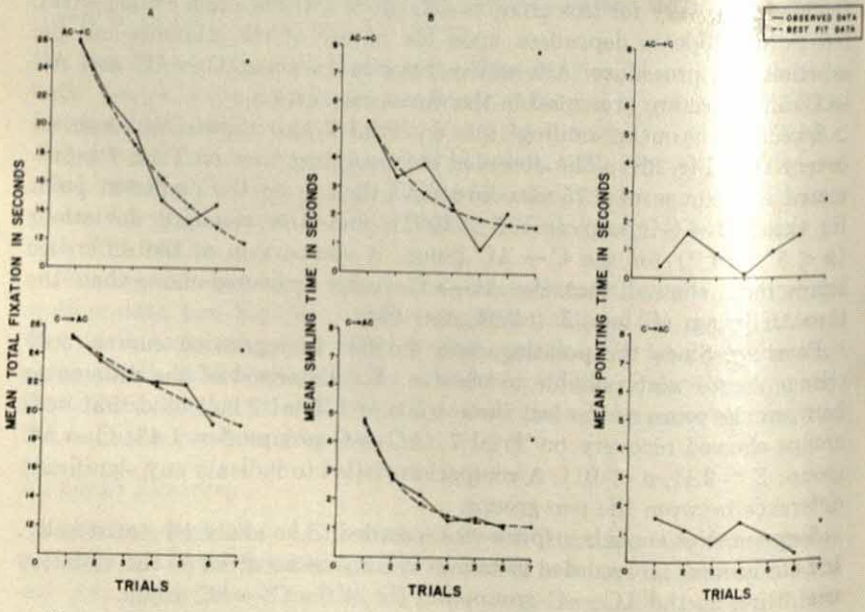


FIG. 3. The mean fixation, smiling, and pointing data and the best fit regression curves for the CL → SL and SL → CL groups.

$5.11e^{-.5x}$ ($F = 69.40$, $p < .005$), and for the SL→CL group was $Y = 1.20 + 2.87e^{-.89x}$ ($F = 11.34$, $p < .05$). There was no significant difference in response decrement for the two groups.

Response Recovery

Fixation. Both the CL→SL and SL→CL groups showed response recovery on the violation trial (see Fig. 3A). The observed mean TF on Trial 7 for the CL→SL group represented a distance of 2.64 standard deviations ($p < .004$) from the point predicted by the regression curve, and for the SL→CL group, a distance of 8.31 standard deviations ($p < 5 \times 10^{-17}$) from the predicted point. A comparison of the difference scores revealed that the recovery for the SL→CL group was greater than that for the CL→SL group (*U* test, $Z = 1.88$, $p < .07$).

Smiling. Only the SL→CL group showed significant recovery of the smiling response on the violation trial (see Fig. 3B). The observed response for the CL→SL group represented a distance of 0.73 standard deviations from the predicted point, and the SL→CL group, a distance of 8.02 standard deviations ($p < 5 \times 10^{-16}$). Like fixation, recovery of the smiling response was greater for the SL→CL group although this difference failed to reach significance (*U* test, $Z = 1.34$).

Pointing. Recovery of the pointing response on the violation trial was evident in the data for both groups (see Fig. 3C). The observed response of the CL→SL group was 1.68 standard deviations ($p < .05$) from the predicted point, and of the SL→CL group, 4.50 standard deviations ($p < 1 \times 10^{-5}$). As for the fixation and smiling data, the recovery for the SL→CL group was greater than for the CL→SL group, but again this difference did not reach significance (*U* test, $Z = 1.67$).

Surprise. There was not enough surprise recorded for any statistical analysis. However, nine instances of surprise were recorded in the CL→SL group, and eight of these were on either Trial 1 or 7, the trials when a new stimulus was presented. For the SL→CL group, all seven of the recorded occurrences of surprise were on the violation trial.

DISCUSSION

Both the AC→C and C→AC groups' decrement data fit negative exponential functions. However, the C→AC group showed less response decrement. The data suggest that decrement differences exist between stimuli differing only in the presence or absence of color. If the addition of color can be considered to increase the intensity or complexity of the stimulus array, these present results agree with other studies indicating habituation differences as a function of complexity in the young child (Ames, 1966; Brennan, Ames, & Moore, 1966; Cohen, 1965).

The recovery data for the two groups were observed; the magnitude of the recovery was greater when the change consisted of the addition of color than when subtraction of color constituted the change. Again, assuming that the addition of color constitutes increased complexity or intensity, the present work with young children is consistent with that of adults showing that more complex or intense stimuli result in greater OR's than less complex ones (Bernstein, 1967; Zimmy & Schwabe, 1966). However, before this can be firmly stated, it is necessary to deal with the problem of the interaction between the amount of response decrement and magnitude of response recovery. Figure 2A indicates that for both groups the response on Trial 7 was essentially equal. It is possible that the amount of recovery has some functional limit (perhaps as a function on the initial response), so that the degree of decrement determines the magnitude of recovery. While this is an important possibility, the smiling data appear to refute this (see Fig. 2B). It is clear that for this response the amounts of decrement are equal for both groups and yet the amount of recovery is significantly different. Parenthetically, the data also demonstrate that recovery is not limited by the magnitude of the initial response.

It is clear that intensity changes can affect the magnitude of an OR in children. However, it may be that intensity change is not the only condition that can affect the OR. In that response decrement and total fixation times were equal for the CL → SL and SL → CL groups, and because the number and color elements were held constant, it is difficult to discuss intensity or complexity differences between these stimuli. While failing to reach statistical significance the data does consistently suggest that a change to curved lines produces a larger OR than a change to straight lines.

It is difficult to state the reasons for the differential OR magnitudes. However, it should not be surprising to discover that some environmental changes produce more immediate attending than do others. James (1890) has suggested a useful conceptionalization when he discusses two types of attending behavior: passive or involuntary and active or voluntary. A stimulus which compels the organism to attend, such as a loud noise, may be said to elicit a passive attending response. Other stimuli, however, elicit attention by means of their relationship to the organism in what James calls their associational value. Not all stimuli have equal associational value and the nature of this value is reflected in the magnitude of the OR. Thus, the cry of a baby and the noise of a moving truck (equal in intensity) should produce unequal OR's to the listening mother.

In order to explore this problem of differential OR's, a recent investigation by Lewis (1969) introduced changes in color, form, number, and

rotation and found a significant hierarchy of response recovery. Demonstrating that this hierarchy was related to other dimensions of cognitive salience, it was argued that this OR paradigm is a useful vehicle for mapping out stimulus salience hierarchies in the young child and from a developmental point of view, their change over time.

REFERENCES

AMES, E. W. Stimulus complexity and age of infants as determinants of the rate of habituation of visual fixation. Paper presented at the Western Psychological Association Meeting, Long Beach, California, April, 1966.

BERNSTEIN, A. S. The orienting reflex and direction of stimulus change. Paper presented at the Eastern Psychological Association Meeting, Boston, April, 1967.

BRENNAN, W., AMES, E. W., & MOORE, R. W. Age differences in infant's attention to patterns of different complexities. *Science,* 1966, 151, 354–356.

COHEN, L. B. Observing responses, visual preferences and habituation to visual stimuli in infants. Unpublished doctoral dissertation, U.C.L.A., 1965.

JAMES, W. *The principles of psychology.* New York: Holt, 1890.

LEE, L. C. Concept utilization in preschool children. *Child Development,* 1965, **36,** 221–227.

LEWIS, M. A developmental study of learning within the first three years of life: Response decrement to a redundant signal. Princeton, N. J.: ETS Research Bulletin 68-62, November, 1968.

LEWIS, M. The meaning of an orienting response: A study in the hierarchical order of attending. Paper presented at Society for Research in Child Development meetings, Symposium on *The Orienting Response: Issues in Developmental Inquiry,* Santa Monica, California, March, 1969.

LEWIS, M., & GOLDBERG, S. The acquisition and violation of expectancy: An experimental paradigm. *Journal of Experimental Child Psychology,* 1969, **7,** 70–80.

LEWIS, M., GOLDBERG, S., & RAUSCH, M. Attention distribution as a function of novelty and familiarity. *Psychonomic Science,* 1967, **7,** 227–228.

SOKOLOV, Y. N. *Perception and the conditioned reflex.* New York: Macmillan (S. W. Waydenfeld, trans.), 1963.

ZIMMY, G. H., & SCHWABE, L. W. Stimulus change and habituation of the orienting response. *Psychophysiology,* 1966, **2,** 103–115.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 306–313 (1969)

Fixed-Ratio and Fixed-Interval Schedules of Cartoon Presentation[1]

MICHAEL D. ZEILER[2] AND CRYSTAL A. KELLEY
University of Iowa

Children turned cartoon movies on briefly by pressing one lever: cartoons appeared according to either fixed-ratio (FR) or fixed-interval (FI) schedules. Presses of a second lever did not present the cartoons. The rate of responding to the FR or FI lever increased, and responding to the other lever decreased. Generally, there were higher and steadier response rates with the FR than with the FI schedules. Further evidence for schedule effects was that children changed to FR after first being trained on FI rapidly attained a high, steady response rate, and children changed to FI after initial training on FR attained a lower, more erratic rate. The duration of each cartoon presentation appeared to be relevant in determining the ability of the cartoons to maintain responding.

In a series of experiments, Baer controlled the behavior of children by having the presence or absence of cartoon movies depend on responses. Responses that avoided the termination of the cartoons increased in frequency (Baer, 1960), and responses that turned the movie off became less probable (Baer, 1961, 1962). On the basis of these data together with the observation that *S*s would respond to turn the film back on (Baer, 1960), Baer concluded that cartoons operate as positive reinforcers for young children.

As of this time, cartoons have not been studied in a common paradigm of positive reinforcement, namely, the simple situation involving their presentation dependent on responses according to intermittent schedules. The present experiment studied whether the opportunity to view cartoons would control the rate and the temporal distribution of the responses that provided each opportunity. Such a situation requires that each presentation be of relatively brief duration, and that movie onset rather than offset be correlated with responding. Information about briefly

[1] This research was supported by Research Grant HD 02845 from the National Institute of Child Health and Human Development and by Undergraduate Research Participation Program Grant GY4356 from the National Science Foundation to the University of Iowa.

[2] Now at Emory University.

viewable movies is not available from Baer's research, because, in his experiments, appropriate behavior would keep the films on continuously.

METHOD

Subjects

The *S*s were preschool children, ten girls and eight boys, ranging in age from 4 to 5 years. The children were studied during the normal school day.

Apparatus

A 28 × 16 × 12-inch wooden box was located on the floor of the dimly lit experimental room, approximately 6 feet in front of an 18 × 24-inch translucent Plexiglas screen (Polacoat). The box contained two response levers made from sponge mop handles (Bijou, 1957) located 8 inch apart.

A 16-mm sound motion picture projector behind the screen showed animated cartoons, and a loudspeaker placed in the experimental room enabled *S* to hear the sound-track. An electrically operated shutter could be interposed between the projector lens and the screen to block the image and prevent the movie from being seen; the shutter operation also disconnected the soundtrack from the speaker. The projector continued to operate during shutter closures, but *S* could neither see nor hear the film.

Typographic material was removed from 12 black and white animated cartoons, and the films were placed on three reels of four cartoons each. The cartoons included five of Woody Woodpecker, two of Dynamo Doc, and one each of Chilly Willy, Mighty Mouse, Mickey Mouse, Daffy Duck, and Inspector Willoughby.

Transistorized programming equipment, electromechanical counters, a running time meter, and a cumulative recorder controlled procedures and recorded data automatically.

Procedure

Each *S* was brought into the experimental room and was seated on the floor in front of the response-lever box. The *E* told the child to "press the levers and see what happens."

All *S*s had a concurrent schedule of cartoon presentation: Responses to one lever turned the films on and responses to the other lever did not. The lever associated with each of the two components of the concurrent schedule alternated on successive days of training; at the start of a session, *E* pointed to the cartoon lever, the one that produced the movies, and said, "Pressing this lever will turn the cartoons on." For one group of nine children (FR group), the schedule was concurrent fixed ratio (FR)

extinction: every *n*th response to the cartoon lever turned the film on. For the other group of nine children (FI group), the schedule was concurrent fixed-interval (FI) extinction: the first response to the cartoon lever after the time specified by the interval elapsed presented the film. Intervals were timed from the end of each viewing period, and only responses occurring when the cartoons were off met the ratio requirements.

In their first session, 12 *Ss* had 10-second periods of cartoon presentation. Thereafter, and throughout the entire experiment for the rest of the children, each viewing period was 30 seconds.

Ratios progressed from FR 1 to FR 30 for the FR group. When *Ss* reached FR 30, they were maintained at that value for at least one and usually for two complete sessions. Three children had an additional session at FR 60, and two had another session with the schedule changed to concurrent FI 30 seconds extinction.

After a progression from FR 1 to FR 5, *Ss* in the FI group were put on a fixed-interval schedule. Schedule values were raised from FI 1 second to FI 30 seconds. Children were maintained for at least one and usually for two complete sessions at FI 30 seconds. Four children had an additional 1 or 2 days at FI 60 seconds, and two had another session with the schedule changed to concurrent FR 30 extinction.

No responses to the cartoon lever turned the film on if they occurred within a specified time after a response to the extinction lever. The time was 3 seconds during the first session for 15 *Ss* and increased to 6 seconds in the first session for the other three children and on all days thereafter. The purpose of this procedure was to prevent rapid alternations between levers from being accidentally correlated with the appearance of the cartoons. Similar temporal contingencies reduced the frequency of switching in a variety of concurrent schedules (Catania, 1966; Findley, 1958).

Sessions lasted for the length of one reel of cartoons (20–23 minutes). Children averaged 4 days between successive sessions and participated for a maximum of six sessions. The children saw each reel in turn on successive days. On all days after the first, the session began with a 30-second presentation of the cartoon.

RESULTS

If the periods of cartoon viewing were reinforcing, responding to the lever that turned the cartoons on should be maintained at a substantial rate. Figures 1 and 2 are the cumulative records of the *Ss* displaying the highest and the lowest rates with FR 30, FR 60, FI 30 seconds, and FI 60 seconds on the last day that the schedule was in force throughout an entire session. Figure 1 shows behavior under fixed-ratio schedules, and Fig. 2 is for the fixed-intervals.

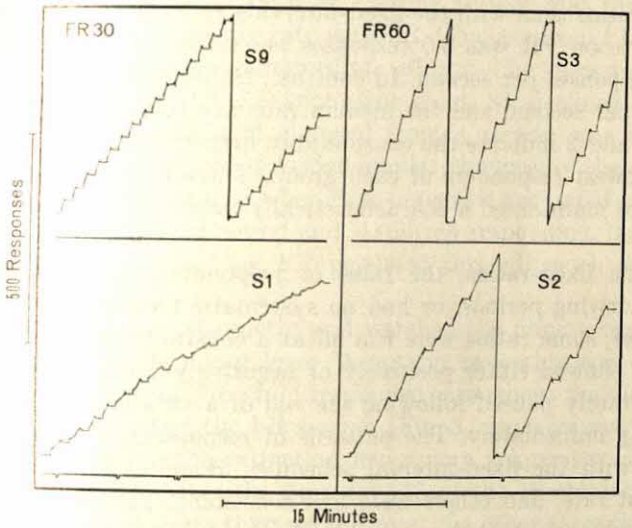


FIG. 1. Cumulative records of fixed-ratio behavior for the last day of exposure to the schedule. The response pen offset during film presentations; slash marks on the record indicate cartoon periods when the pen did not offset. Deflections of the event pen indicate the occurrence of response to the extinction lever; the pen remained deflected for the 6-second period in which responses to the other lever could not turn on the cartoon.

Both ratio and interval schedules maintained lever pressing at all of the parameter values studied but there did appear to be effects of particular schedules. Although there was overlap between the *S*s of each group, a Mann-Whitney *U* test showed that the response rates were higher with

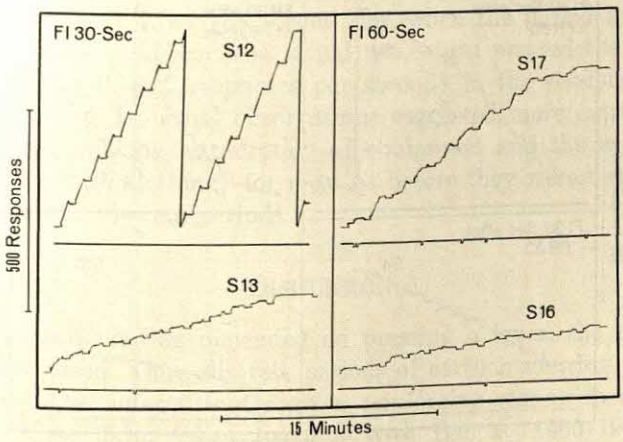


FIG. 2. Cumulative records of fixed-interval behavior for the last day of exposure to the schedule. Recording was as in Fig. 1.

the fixed-ratios than with the fixed-intervals ($p < .05$, 2-tailed test). The highest rate on FR was 3.0 responses per second, and the median rate was 1.5 responses per second. In contrast, the highest rate on FI was 2.1 responses per second, and the median rate was 0.4 responses per second. Figures 1 and 2 indicate the considerably higher rates on FR than on FI for the slowest responders of each group. These data suggested that the fixed-ratios maintained a characteristically faster rate than did the fixed-intervals.

With the fixed-ratios, the rates of responding were either constant between viewing periods, or had no systematic temporal pattern. In the latter cases, some ratios were run off at a constant rate, but others were erratic or showed either positively or negatively accelerated responding. Children rarely paused following the end of a viewing period, but began responding immediately. The patterns of responding were not markedly different with the fixed-interval schedules. Some children responded at a constant rate, and others were less consistent. There were more cases of positively accelerated responding between viewing periods with the fixed-intervals than with the fixed-ratios, and more variability within- and between-*S*s with the fixed-interval.

Figure 3 shows the behavior of children switched from concurrent FR extinction to concurrent FI extinction and vice versa. Those changed from FI 60 seconds to FR 30 showed a marked increase in rate: The rate of *S*14 increased from .3 responses per second on FI to 2.0 responses per second on FR, and the rate of *S*15 rose markedly above its previous level after the fourth cartoon presentation on the FR schedule. The previously

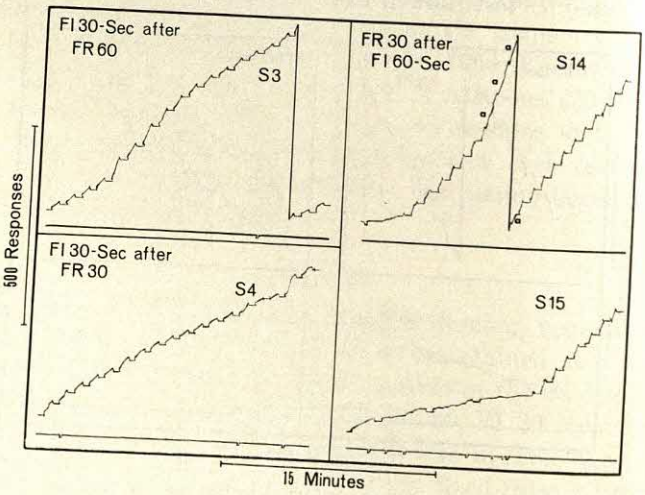


FIG. 3. Cumulative records for fixed-ratio behavior following days with fixed-interval schedules and vice versa. Recording was in Fig. 1.

erratic pattern of responding between viewing periods that characterized FI behavior shifted to a steady rate with FR. Changes from FR to FI did not have equally obvious and immediate effects: The rates remained at the FR level for much of the session. One child, *S*3, declined in rate by the middle of the session and changed from a steady rate to a more erratic pattern; the other, *S*4, revealed similar changes by the end of the day. These data indicated that when fixed-intervals generated a moderate or low rate, fixed-ratios increased and stabilized responding, but that the high steady rate generated by FR persisted through much of a single day of FI.

The children usually sat quietly and watched the films when they appeared without pressing either lever. A notable exception was *S*14 when switched from FI to FR. This child responded throughout the film presentation in the early part of the FR session (Fig. 3, segments marked "a").

Each *S* responded to the extinction lever more frequently in the first than in later sessions. After the first day, 15 of the 18 children pressed the extinction lever no more than seven times; the other three *S*s behaved similarly by the third session. The slowest children to eliminate responding to the extinction lever all began training with the short (3-second) delay period between extinction responses and a film presentation for a response to the other lever.

Four children were not observed on any schedule value for an entire session. One left school due to family vacation, and the other three, two trained on FR and one on FI, refused to return after the first session. These three children all had received 10-second cartoon presentations. No *S* started with 30-second periods refused to return, and one child, who was reluctant to return after a first day with 10-second cartoon durations, participated readily after the second day when the period increased to 30 seconds. Those children who would not return pressed the lever at a low rate (about 0.1–0.2 responses per second) in the sessions in which they participated. Informal observations suggested more extraneous behavior (thumb-sucking, exploration of equipment and the experimental room, playing with clothing) for most *S*s before they were changed from 10- to 30-second viewing periods.

DISCUSSION

When cartoon viewing depended on pressing a lever, the rate of responding increased. Thus, discrete periods of cartoon viewing functioned similarly to other intermittently given reinforcing stimuli in controlling behavior. These data, taken together with Baer's (1960–1962), show that children will respond to produce cartoons briefly, will avoid the termination of cartoons, and will stop performing responses that termi-

nate a cartoon. Cartoons, therefore, can control the behavior of many children in a variety of reinforcement paradigms.

The present data do suggest that the duration of each viewing may determine the effectiveness of cartoons. With periods that are too short, cartoons may be ineffective, and based on some children's refusal to return to the situation, may even be avoided. Whether duration of presentation alone is the major variable, or whether duration interacts with particular schedules to determine the functional properties of cartoons, is unknown. In the light of Kelleher and Morse's (1968) data showing that electric shock—a stimulus commonly used in studies of avoidance, escape, and punishment—will increase the probability of a response producing it when the shock is programmed at certain intensities and according to particular schedules, it would not be surprising if cartoons either increase or decrease responding depending on such parameters as duration and schedule of presentation.

The behavior observed with fixed-ratio and fixed-interval schedules was like that reported by Long, Hammack, May, and Campbell (1958) with trinket, penny, and photographic slide reinforcement. That study also found little postreinforcement pausing with schedules up to FR 60, generally higher response rates with fixed-ratios than with fixed-intervals, and more variability with the fixed-interval schedules. The Long *et al.* data suggested that the fixed-interval behavior does become somewhat more stable if a large number of sessions are conducted and if the power of the reinforcing stimulus can be maintained.

The present study allows no conclusion as to whether the effective event was the presentation of the cartoons or was the termination of periods without cartoons. Long *et al.* (1958) recognized the possibility of escape from the absence of stimulation and attainment of a stimulus operating jointly and speculated that a combination of acquisition of a stimulus and escape from an absence of stimulation may be a general phenomenon extending beyond work with children: " . . . the acceleration to a terminal rate reported for lower organisms responding for homeostatic reinforcers may be in part for positive consequences and in part for the removal of an aversive one" (Long *et al.*, 1958, p. 332). Recent data (Morse and Kelleher, 1966) show that the termination of noxious stimuli produces patterns of responding very similar to those observed with food presentation (Ferster and Skinner, 1957). Since escape and stimulus presentation affect behavior similarly, separating their influences would appear to be difficult. Evaluation of the possibility that acquisition and escape may operate together awaits techniques that can isolate the effects of terminating the absence of cartoons or other stimuli from the effects of the presentation of the event.

REFERENCES

BAER, D. M. Escape and avoidance response of pre-school children to two schedules of reinforcement withdrawal. *Journal of the Experimental Analysis of Behavior,* 1960, 3, 155–159.

BAER, D. M. Effects of withdrawal of positive reinforcement on an extinguishing response in young children. *Child Development,* 1961, **32,** 67–74.

BAER, D. M. Laboratory control of thumbsucking by withdrawal and representation of reinforcement. *Journal of the Experimental Analysis of Behavior,* 1962, **5,** 525–528.

BIJOU, S. W. Methodology for an experimental analysis of child behavior. *Psychological Reports,* 1957, **3,** 243–250.

CATANIA, A. C. Concurrent operants. In W. K. Honig (Ed.), *Operant behavior: areas of research and application,* pp. 213–270. New York: Appleton, 1966.

FERSTER, C. B. AND SKINNER, B. F. *Schedules of reinforcement.* New York: Appleton, 1957.

FINDLEY, J. D. Preference and switching under concurrent scheduling. *Journal of the Experimental Ananysis of Behavior,* 1958, **1,** 123–144.

KELLEHER, R. T. AND MORSE, W. H. Schedules using noxious stimuli. III. Responding maintained with response-produced electric shocks. *Journal of the Experimental Analysis of Behavior,* 1968, **11,** 819–838.

LONG, E. R., HAMMACK, J. T., MAY, F., AND CAMPBELL, B. J. Intermittent reinforcement of operant behavior in children. *Journal of the Experimental Analysis of Behavior,* 1968, **11,** 819–838.

MORSE, W. H. AND KELLEHER, R. T. Schedules using noxious stimuli. I. Multiple fixed-ratio and fixed-interval termination of schedule complexes. *Journal of the Experimental Analysis of Behavior,* 1966, **9,** 267–290.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 314–327 (1969)

An Experimental Analysis of the Criteria Used by Children to Judge Quantities

GRAEME S. HALFORD
The University of Newcastle

This study was an attempt to investigate the ability of various groups of children to classify containers which were unequal to an original container. Material was poured from an original container (O) into a new container (I) of the same quantity but different height and breadth. A series of standard containers equal to I in diameter but either shorter or taller was placed between O and I. The S was asked whether the standards would overfill or underfill O. In a control condition *Ss* did not know that O and I were equal in quantity. Not only *Ss* with conservation but also nonconservation *Ss* performed better in the experimental condition than in the control condition, provided the *Ss* were over 5 years of age. *Ss* under 5 years of age showed no difference between experimental and control treatments. The results were interpreted as showing that as from the age of 5 years, *Ss* construct a classifiactory system which provides logical criteria for judging the quantity of containers. This classifiactory system corresponds to the conjunction of heights and breadths.

The conservation of quantity is described by Piaget (1950, 1952, 1957) in terms of the logical multiplication of heights and breadths of containers. Suppose material is poured from an original container O (which may be short and wide) into another container of *identical size* I, which may be taller and narrower. In general, whenever there is a series of equal containers (of the same shape) differing in dimensions, for any value of one dimension there is one and only one of the other dimension.

Recognition of equality of quantity implies recognition of an implicit classificatory system of containers. Consider containers which are all the same breadth as I above. Recognition that I is equal to O implies that any container shorter than I is less than O, and any container taller than I is larger than O. Furthermore, this is precisely equivalent to saying that with any breadth, the height of I is the only height which a container may have which is equal to O. This point is very important. Suppose *Ss* recognize that, where containers are equal to O, then for any breadth there may be only one height. This recognition may depend on the ability of *Ss* to classify containers which do not have this height as unequal to O. For instance, if there is a container SI (equal to I in breadth but shorter)

then SI is necessarily less than *O*. Similarly a container taller than *I* but equal to it in breadth is necessarily larger than *O*. Thus, the classification of these containers implies that, given the breadth of SI, *I*, TI, etc., *I* alone has the height which will make it equal to *O*. Thus, understanding of the relations between equal containers implies a classificatory system which must deal with unequal containers as well.

This classificatory system corresponds to part of the classificatory system implied by conjunction. Any logical operation has a truth table, which classifies events into mutually exclusive and exhaustive categories, and assigns to each category a certain truth value. A classificatory system for containers may be constructed which corresponds to the truth table for conjunction. This is shown in Table 1.

TABLE 1

CLASSIFICATORY SYSTEM FOR ANY CONTAINERS

Material is poured from *O* to *I* or one of the containers in the column beneath it (narrower or wider than *I*, NI, and WI: shorter or taller than *I*, SI, and TI). Any container may be classified according to whether it has the height, breadth, both or neither of *I*, which is equal to *O*. ($\overline{h_x}$ = "not the same height as h_x"). The truth table for conjunction is also shown using the purely formal symbols A and B.

Original	Identity	Height	Breadth	Classification	A	B	A · B
O	*I*	h_x	b_x	Equal	1	1	1
	NI, WI	h_x	$\overline{b_x}$	Unequal	1	0	0
	SI, TI	$\overline{h_x}$	b_x	Unequal	0	1	0
	Miscellaneous	$\overline{h_x}$	$\overline{b_x}$	Indeterminate	0	0	0

If, as Piaget (1950, 1957) claims, conservation depends on logical multiplication of height and breadth, then a *S* who understands conservation must have an implicit classification system isomorphic to that of Table 1. The logical operation itself cannot be directly investigated empirically. The existence of the classificatory system may be investigated. It is the purpose of this work to investigate the ability of children to classify containers in category three of Table 1. The procedure is essentially similar to that of an earlier investigation (Halford, 1968).

Children may recognize the identity of the material in *O* and *I* without conserving, since identity is a necessary but not a sufficient condition of conservation (Bruner, Olver, and Greenfield, 1966). Identity as such is merely an observed fact. It means that the material in *I* was formerly in *O*, but this has no consequences for its quantity. However, once a classificatory system such as that of Table 1 has been constructed, *identity* does have consequences for the dimensions of the containers. Identity

becomes integrated into a system which links it to observations of and inferences about dimensions, and this integration may well be essential to the achievement of conservation.

Children who have conservation should have the classificatory system corresponding to heights and breadths. Such *S*s should be able to classify containers appropriately in any of the categories shown in Table 1. If *S*s can classify containers in category three of Table 1, they must be capable of recognizing that such containers are different from *O* because they have the same breadth as *I* but a different height. Since they can observe whether the height is less than or greater than that of *I*, they should be able to state whether such containers are less than or greater than *O*.

Children who do not have conservation should not have this classificatory system. They should therefore be incapable of recognizing that containers SI— — to SI are smaller than *O* because they are shorter than *I*, and that containers TI to TI++ are larger than *O* because they are taller than *I*. In short, they should be incapable of correctly responding to containers which are the same diameter as *I*, but different in height, beyond the level of performance which is possible by direct perceptual judgment.

EXPERIMENT I

Method

Subjects[1]

The *S*s were 208 school children aged 6–8 years from four state schools in the Newcastle area. Those *S*s offered by teachers were accepted and tested unless they had to be discarded. The above sample of 208 *S*s does not include 19 *S*s discarded for various failures to comply with procedural requirements. There were four experimental groups of 52 *S*s and four *E*s. Each *E* tested 52 *S*s, 13 in each experimental group. Each *E* was assigned to a different school.

Apparatus

The apparatus consisted of 16 cylindrical, clear perspex beakers similar to those described earlier (Halford, 1968). The six *originals* were equal in quantity, and had dimensions (internal diameter, internal height, in inches) as follows: IE, 1.25×9.23; 1C, 1.50×6.60; 1B, 1.75×4.75;

[1]In carrying out these experiments, generous co-operation was received from the New South Wales Department of Education, from several schools in the Newcastle area, from the Maitland Nursery School, from "The Hill" Kindergarten, and from Mr. R. Wilson, Misses A. Muir, D. Peel, and P. Huby, and Mrs. J. Shipley.

1A, 2.25 × 2.85; 1D, 2.75 × 1.90; 1F, 3.25 × 1.38. The six standards all had the same breadth (2.00 inches) and their heights were: SI——, 2.85; SI—, 3.15; SI, 3.45; TI, 3.75; TI+, 4.05; TI++, 4.35. The identity container *I* was 2.00 wide by 3.60 high. Thus, there were three standards shorter than the identity (SI—— to SI) and three taller (TI to TI++). Three of the originals also were shorter than *I*, (1A, 1D, and 1F) and three were taller (1E, 1C, and 1B).

There were three containers used solely for the conservation of quantity pretest. These were: 2A, 2.75 × 3.00 (2 of) and 2D, 1.75 × 7.45.

Procedure

All *S*s were first given the classical test for conservation of quantity, using the procedure described previously (Halford, 1968).

The testing procedure for the experimental or known *quantities* (K) groups was as follows. The *S* and *E* were seated on adjacent sides of a table. One of the originals, already levelfilled, was placed a few inches in front of *S*. The identity container *I* was placed *empty* a few inches to the side of *O*, and *O* was poured into *I*. A standard container, already levelfilled, was placed between *O* and *I*. The *E* said:

"If I pour this one (standard) into here (original) will it overflow, or will there not be enough?"

The question was repeated with appropriate additional comments when necessary to achieve understanding. After *S* answered, the standard was poured into *O*, and *S* was asked "were you right?". If this question was not correctly answered, explanation was given. The procedure of pouring the standard into *O* was discontinued when *S* clearly understood what he was being asked to do.

The procedure for the control or unknown quantities (U) condition was the same except that *O* was emptied into a bowl nearby, and *I* was already levelfilled. Thus, U *S*s did not see *O* poured into *I*, and did not therefore know that these two containers were equal in quantity. The procedure forced *S*s to judge the standards as either greater or less than the original.

Each of the six originals were presented with each of the six standards in this way. The 36 paired comparisons were presented in a pre-arranged random order, 12 in each of three 15-minute interviews.

At the conclusion of the third interview, the *S*s were given the classical test for conservation as a posttest.

Experimental Design and Hypotheses

On the basis of the classical test for conservation, the *S*s were assigned to a Conservation (C) or a Nonconservation (N) group. Half of each

group was then randomly allocated to the K or U experimental treatment groups. There were, thus, four groups of 52 *Ss* each.

Conservation *Ss* should be able to classify the test containers relative to *I*. All the test containers have the same diameter as *I*. Therefore conservation *Ss* should recognize that *I* alone has the correct height to fill *O*. Thus, *I*, may be used as a criterion for deciding whether a test container is less than or greater than *O*. If conservation implies the existence of the classificatory system discussed earlier, C *Ss* should perform this task, whereas N *Ss* should be unable to do so. Thus, N *Ss* should have to rely on direct comparison of standards and *O*, and make their judgments on perceptual criteria alone. Since the differences to be discriminated are small, use of *I* as a criterion of quantity should result in improved performance for C *Ss*. Thus, C *Ss* should perform better in the K condition than in the U condition. The N *Ss* should show no such difference between K and U treatment conditions.

Thus, the key hypothesis is that there should be more difference between K and U groups in the case of C *Ss* than in the case of N *Ss*.

Results

The mean number of correct responses out of 36 are shown for each experimental group and each *E* in Table 2. Analysis of variance of these data showed that all three main effects are significant at .001 level or beyond. For K/U, $F_{1/192} = 38.7$; for C/N, $F_{1/192} = 12.1$; for experimenters-schools, $F_{3/192} = 11.4$. None of the *F*s for interactions involving *E*s is significant. This indicates that the procedures are reliable across *E*s.

TABLE 2
MEAN NUMBER OF CORRECT RESPONSES (out of 36) FOR EACH EXPERIMENTAL GROUP AND EACH EXPERIMENTER

Experimenter schools	CK	CU	NK	NU
A	29.2	26.0	29.3	23.1
B	27.5	21.9	24.0	20.2
C	27.2	23.8	24.0	22.5
D	25.5	20.9	21.6	21.2
Group means	27.4	23.2	24.7	21.8

The main hypothesis was that there should be more difference between K and U groups in the case of C *Ss* than in the case of N *Ss*. This corresponds to a planned comparison which gives weights of $+\frac{1}{2}$ to CK and NU and weights of $-\frac{1}{2}$ to NK and CU groups. This comparison yields $F_{1/129} = .87$, which is not significant. (The other *F* tests are equivalent to

corresponding planned comparisons, but are more conveniently stated as ominbus F tests.)

Comparisons among all pairs of means for the four groups CK, CU, NK, and NU were based on a standard error of the difference between means of .81. $df = 102$. All differences are significant at the .05 level or beyond except that between groups CU and NK and that between CU and NU. These last are significant by one-way test only.

Of the original 104 N group *Ss*, 41 changed to conservation on the posttest. Of the 104 C group *Ss*, 14 changed to nonconservation on the posttest. On the null hypothesis that there will be an equal number of conservation and nonconservation *Ss* on the posttest (i.e., an equal number of changes in each direction) $X^2 = 13.50$, $df = 1$, $p < .001$. Thus, it can be concluded that significantly more *Ss* change from nonconservation to conservation than the reverse during the period of the experiment.

Of the 41 N *Ss* who changed to conservation, 23 were in the K group. On the null hypothesis that conservation performance on the posttest is independent of membership of K or U treatment groups, $X^2 = .64$, $df = 1$, $p >> .05$. If the same test is used for all *Ss*, whether originally in the C or N groups, $X^2 = 2.06$, $df = 1$, $.10 < p < .20$. (Parenthetically, if all *Ss* who change from C to N or N to C on the posttest are eliminated, the mean correct responses out of 36 are, for group CK, 27.10; group CU 23.10; group KU, 24.69, group NU, 21.53. Comparing these with the column averages in Table 2, it is apparent that the elimination of *Ss* who change conservation status makes no difference to the relationship between the experimental groups.)

DISCUSSION

The only way in which the results of this experiment are in accordance with prediction is that the C *Ss* do perform better in the K condition than in the U condition. However, this is also true of N *Ss*, and the difference is not significantly greater in the case of C *Ss* than in the case of N *Ss*. This means that N *Ss* must also have the skill attributed to conservation *Ss* in the introduction. Thus, N *Ss* must be capable of judging quantity on logical as well as on perceptual criteria.

Referring to Table 1 in the introduction, it is apparent that N *Ss* must have available at least part of the classificatory system which corresponds to conjunction of height and breadth. This has several implications. The first is that these *Ss* have apparently *partially* constructed the concept of conservation, since they can judge quantity on logical criteria but do not conserve quantity. This is consistent with the view that conservation is acquired gradually.

Second, it means the classificatory system is apparently constructed *before* conservation occurs. This would be consistent with the proposition

that conservation *results from* the classificatory system, rather than the reverse. It is incompatible with an interpretation in terms of transitivity. It means that *S*s who apparently do not know (at least explicitly) that *O* is equal to *I*, still know that SI is less than *O* and TI is larger than *O*. To apply transitivity, *S* must know that $O = I$. It is, however, quite compatible with the view that conservation is attained by gradually learning to classify containers relative to *I*. In other words, the *S* may learn that $I = O$ by learning that SI $<$ O and TI $>$ O.

The other finding from the experiment is that C *S*s are superior to N *S*s in both the K and U conditions. This finding replicates the findings of an earlier study (Halford, 1968). The superiority in the U condition means that C *S*s are superior to N *S*s even where judgments are on a perceptual basis. This means that both groups of *S*s can judge perceptually the change in height needed to accompany a given change in breadth where quantity remains constant, but C *S*s are superior in making this judgment. This ability itself cannot lead to logical recognition of conservation. This is not to say, however, that it is not *conducive* to construction of the conservation concept.

Having found that a logical criterion of quantity can be utilized before conservation is attained by the classical test, it is now important to establish the earliest age at which this criterion can be applied. Kendler and Kendler (1962) have observed a change in cognitive behavior which occurs at age 5. Also, McLaughlin (1963) has suggested that the underlying predisposition to acquire concrete logical operations might be a memory span of four items, which occurs at 4½ to 5 years of age. Thus, it is possible that the classificatory system shown in Table 1, which enables *S* to utilize a logical criterion of quantity, may first appear after about 5 years of age.

EXPERIMENT II

In Exp. II, a sample of *S*s over 5 years, but who have not yet attained conservation, will be compared with a sample of pre-school *S*s, using the same procedure as in Exp. I. It is expected that the 5 year olds will perform better in the K condition than in the U condition, whereas the preschool *S*s will not.

Method

The procedure was as that for Exp. I. The 36 paired comparisons of beakers were interspersed with 16 memory-span items, eight of three digits each and eight of four digits each. The digits were selected from random permutations of decimal digits and spoken to *S* at the rate of one per second, *S* being instructed to repeat them.

Subjects

The *Ss* were 60 children aged between 60 and 72 months from two schools in the Newcastle area, and 60 pre-school *Ss* aged 41–60 months from two Newcastle nursery schools. Half of each group were tested in the K condition and half in the U condition. One *E* tested 19 *Ss* in each of the four experimental groups, and one tested 11.

All *Ss* were pretested for conservation of quantity, and eliminated if they were found to have conservation. Only five *Ss* were eliminated for this reason. *Ss* were accepted in the order provided by the school. Only three *Ss* were eliminated for failure to meet procedural requirements.

There were four exprimental groups comprising 5 year olds in the K condition (FK), and the U condition (FU) and pre-school *Ss* in the K and U condition (PU and PK).

Results

The mean number of correct responses out of 36 for each of the four experimental groups is shown in Table 3. Analysis of variance of these data yields a significant main effect for age ($F_{1/112} = 5.36$, $p < .05$), but not for the experimental treatment ($F_{1/112} = 2.01$, $p > .05$). The effect of experimenters was also non-significant ($F_{1/112} = 1.67$, $p > .05$). All interactions involving experimenters yielded F ratios of less than one, confirming the reliability of effects across experimenters.

The prediction that there will be more difference between the U and K conditions for the 5-year-old (F) group than for the pre-school (P) group corresponds to a planned comparison with weights of $+\frac{1}{2}$ to groups FK and PU, and $-\frac{1}{2}$ to FU and PK. This comparison yields $F_{1/112} = 4.08$, $p < .05$. Thus, this prediction is confirmed.

Comparisons among all pairs of means may be made using a standard error of the differences between means of .86, $df = 58$. This would mean

TABLE 3

PERFORMANCE OF *Ss* FROM BOTH EXPERIMENTS ON 36 JUDGMENTS OF QUANTITY

	Mean correct responses		Percent better than chance			Mean age
	Exp.	Contr.	Exp.	Contr.	E C	
Conservation *Ss*	27.35	23.15	51.9	28.6	52,52	85.1
Preconservation *Ss*	24.73	21.75	37.4	20.8	52,52	79.0
Five-year-old *Ss*	23.33	21.37	29.4	19.4	30,30	67.8
Pre-school *Ss*	20.83	21.33	15.6	18.3	30,30	52.6

that the FK group differs significantly from the other three, which do not differ among themselves.

The number of correct responses made by the P group *Ss* (20.83 is significantly more than the 18 expected by chance ($t = 4.64$, $df = 29$, $p < .01$)). Thus, even pre-school *Ss* are capable of judging the change in one dimension needed to compensate a a change in the other dimension where quantity is constant. Since they evidently cannot utilize logical criteria of quantity, this must be an "empirical" judgment based on experience, rather than on logical inference.

There were very few errors on the memory span tasks with three items per series. With four items per series, 38 *Ss* over five gave correct responses to seven or eight of the eight series, and 28 P *Ss* did so, $X^2 = 2.73$, $.05 < p < .10$. There is no significant association within groups between memory span performance and number of CRs on the quantities comparison task.

On the posttest, the numbers of *Ss* changing from a nonconservation to a conservation response were; for the FK group 1; FU group, 1; PK group 2; PU group 3. Thus, there is no noticeable tendency for *Ss* of either 53 or 68 months to change to conservation as a result of this procedure. However, there are significantly more changes from N to C by the N group *Ss* in Exp. I than by the F *Ss* in Exp. II. The frequencies are 41 out of 104 and 2 out of 60, respectively, $X^2 = 23.78$, $df = 1$, $p << .001$. The ages are 79 and 68 months, respectively.

Supplementary Analysis

The 36 quantity comparisons may be divided into four groups of nine each, according to whether the originals used were short (1F, 1D, 1A) or tall (1B, 1C, 1E) and whether standards were short (SI——, SI—, SI) or tall (TI, TI+, TI++). The mean number of correct responses out of nine for all experimental groups are shown in Table 4.

A question which is of interest is whether correct responses are spread evenly over the four different kinds of comparisons. An alternative is that *Ss* would find some types of comparisons harder than others. There are three possibilities here: (a) Comparisons may vary in difficulty according to height of original (factor O); (b) comparisons may vary in difficulty according to height of standards (factor S); (c) there may be an interaction between these two factors (O × S).

In the case of (c) if the original is short (S.) a short standard (.S) should appear more likely to overfill it than if the original is tall. There should thus be more errors on cases SS (short original, short standard) than would be expected from frequencies in the table as a whole. Similarly, if the original is tall, a tall standard is more likely to appear to

TABLE 4

MEAN CORRECT RESPONSES ON NINE QUANTITY COMPARISONS FOR FOUR GROUPS IN EXP. I AND II

(Also number of *Ss* who obtained more than 6 [Exp. I] or 5 [Exp. II] CRs)

	Experiment I standards				
	Short		Tall		
	Means	No. of *Ss*[a] succeeding	Means	No. of *Ss*[a] succeeding	Experimental groups
Tall	7.92	48	5.40	32	CK
	8.27	50	2.56	9	CU
	7.56	44	4.63	23	NK
	8.38	49	1.63	6	NU
Orginals Short	7.15	40	6.79	38	CK
	7.27	31	6.10	36	CU
	7.62	27	6.56	39	NK
	4.87	24	6.77	40	NU

	Experiment II standards				
	Short		Tall		
	Means	No. of *Ss*[b] succeeding	Means	No. of *Ss*[b] succeeding	Experimental groups
Tall	7.30	28	4.03	15	>5, K
	8.00	28	2.30	4	>5, U
	6.30	23	3.87	12	<5, K
	6.80	26	3.83	12	<5, U
Originals Short	5.17	19	6.80	24	>5, K
	4.90	17	6.33	21	>5, U
	4.27	12	6.37	25	<5, K
	4.60	12	6.10	21	<5, U

[a] No. of *Ss* who obtained 6 or more CRs out of 9.

[b] No. of *Ss* who obtained 5 or more CRs out of 9.

underfill it than if the standard is short. Thus, again there should be more errors on cases TT than would be expected from the frequencies in the table as a whole.

To test these predictions, the *Ss* were classified according to whether they succeeded or failed on the nine comparisons of each type SS, ST, TS, TT. In Exp. I, the mean number of correct responses overall (to nearest integer) was 6, so *Ss* were regarded as succeeding if they obtained

six or more CRs for each type of comparison. In Exp. II, the mean number of correct responses overall (to nearest integer) was 5, so Ss were regarded as succeeding if they obtained five or more CRs for each type of comparison. The number of Ss succeeding or failing in each type of comparison is given in Table 4.

The factor succeeding or failing may be called P. According to the method described by Sutcliffe (1957), the three effects (a) to (c) above may be analysed by computing X^2 values as follows (a) The effect of O corresponds to interaction O × P; (b) the effect of factor S corresponds to interaction S × P; (c) the effect of O × S corresponds to interaction O × S × P.

The X^2 values corresponding to each of these effects for each experimental group in each experiment are shown in Table 5. In terms of the above prediction O × S × P is the effect of interest. For every group in both experiments, the X^2 is significant at the .02 level or better. (All X^2 except that for group CK in Exp. I are significant at .001 level.) Thus, it may be concluded that errors tend to occur more frequently in SS or TT comparisons than otherwise. Since the number of cases in each experimental group in each of the experiments is equal, the X^2 value may be used as an index of the size of the effect. It is clear that in experiment one this bias is least for group CK, which gives the best overall performance. It is greatest for group NU, which gives the poorest overall performance, and it is intermediate for the other two groups. In the

TABLE 5
CHI-SQUARE VALUES FOR DATA IN TABLE 4
(For explanation, see text)[a]

	Chi-square O × S × P			
	Exp. II		Exp. I	
	>5	<5	C	N
K (Exp.)	22.99	14.27	5.86	20.34
U (Contr.)	16.59	15.35	17.89	47.31
	Chi-square S × P			
	>5	<5	C	N
K	.714	.16	.026	.902
U	12.38	.54	25.66	13.28
	Chi-square O × P			
	>5	<5	C	N
K	0	.16	.03	0
U	.406	.54	.99	1.26

[a] The degrees of freedom are one in every case.

case of Exp. II there are relatively small differences between the sizes of the X^2 values among the different groups.

None of the X^2 values corresponding to the effect of short vs. tall originals approaches significance. In any case, such an effect would have to be interpreted subject to the specific containers used. Nevertheless the X^2 values for the effect of short vs. tall standards are of some interest in that only three are significant (all at .001 level) and these are all for control groups. The data from the main test indicates that different processes may be involved for the pre-school *Ss*, since the effects found in the other cases do not occur in their case. The same is true for this effect. The only nonsignificant X^2 for a control group is for pre-school *Ss*. In all other cases the effect is that tall standards produce more errors for control *Ss* than do short standards.

At present this can only be noted as a *post hoc* observation the significance of which is purely suggestive. It does seem to imply, however, that the performance of control *Ss* differ qualitatively from those of experimental *Ss* except in the case of pre-school *Ss*.

DISCUSSION

The results of this experiment are entirely in accordance with the prediction stated at the end of discussion of Exp. I. The F group *Ss* do perform significantly better in the K condition than in the U condition. This confirms the finding of Exp. I that nonconservation *Ss* may utilize a logical criterion of quantity. However, P *Ss* actually perform better (but not significantly) in the U condition than in the K condition. Thus, the kind of effect sought in Exp. I has been obtained in this experiment, using *Ss* at an earlier stage of development.

Pre-school *Ss* do not utilize the logical criterion of quantity available to *Ss* in the K treatment groups. This implies that they have not begun to construct the logical multiplication of height and breadth of containers, shown in Table 1, and suggests that they cannot classify containers relative to the identity container.

It has been reported many times that even nonconservation *Ss* can estimate the level which material would reach when poured into a second container which is (say) taller and narrower than the original. (Cohen, 1967; Fraisse and Piaget, 1963, Piaget, 1952). These studies confirm this finding, even for pre-school *Ss*. The major difference between the present experiments and those mentioned above is in the provision which is made for a distinction between logical and intuitive judgments. For a logical judgment, any departure from perfect compensation of dimensions necessarily means inequality. If the identity container *I* holds the same ma-

terial which was in the original container, then any container which is the same breadth as *I* but shorter (SI) or taller (TI) must be of different quantity from *O*, however small the departures are. All *S*s in the over 5-year groups have performed better where they were able to use the identity of *O* and *I* as a cue, whereas pre-school *S*s did not. Since they recognize the utility of *I* as a standard, their judgments can be said to be more logical than judgments based on purely intuitive estimation.

In a previous study (Halford, 1968), higher mean scores were obtained in the experimental condition than in the control condition for both conservation and nonconservation *S*s, but the effects were not significant. In the present study these differences clearly are significant for all but the pre-school *S*s. Thus, the prediction made in the first study, and in the introduction to the present study, that only conservation *S*s would show superior performance in the experimental condition is not confirmed. Thus, the crucial difference sought is found, not between conservation and nonconservation *S*s, but between 5-year-old and pre-school *S*s. The present study agrees with the earlier one (Halford, 1968) in showing that conservation *S*s have superior scores to nonconservation *S*s in both the experimental and control conditions.

The association between utilization of a logical criterion of quantity and memory span is not significant in this experiment ($p < .10$). However, the fact that the use of the logical criterion first appears at age 5 suggests that a precondition may be a memory span of four items which is first attained on average at around 4½ to 5 years. In general, McLaughlin (1963) has postulated that the conceptual level which a *S* may reach depends on the memory span.

These experiments were not primarily intended to investigate the acquisition of conservation of quantity. There is evidence that *S*s change to conservation during the experiment, although there is no control for change due to extra-experimental influences during the same period. The period is, however, very short (about 3–4 weeks at most), so the spontaneous rate of change would be expected to be very low. The findings are then consistent with the proposition that this procedure may be effective in inducing conservation of quantity. In this respect it should be noted that no attempt is made to teach *S*s specifically that *O* and *I* are equal in quantity. If this procedure does result in conservation, it is apparently due to some sort of general restructuring, such as the construction (or completion) of a classificatory system.

The number of CRs out of 36 obtained by all groups in both experiments are shown in Table 3 along with the ages of the groups, so that overall comparisons may be made. There is a continuous increase in performance from age 3–4 years to the conceptual stage at around 7 years.

The difference between experimental and control *Ss* widens steadily as age increases. This observation, combined with the observation that non-conservation *Ss* over 5 years can utilize a logical criterion of quantity, suggests that the processes which lead to conservation develop gradually.

These experiments have not shown that conservation is attained at an earlier age than Piaget's tests would indicate. The demonstration that *Ss* utilize the logical criterion of quantity which is available to them does not mean that they recognize the material in *O* and *I* to be *necessarily* the same in quantity, which is what is implied by conservation.

The experiment supports the theory of Piaget insofar as it shows that there is a developmental trend from the use of a logical criteria of quantity (by pre-school *Ss*) to the use of logical criteria (by *Ss* over 5 years). Thus, it supports the view that conservation is based on a logical construction on the part of the child. It goes beyond Piaget's work, however, in that it indicates the possible origins of this construction in the preconservation child.

REFERENCES

Bruner, J. S., Olver, Rose R., and Greenfield, Patricia M. *Studies in cognitive growth*. New York: Wiley, 1966.

Cohen, G. M. Conservation of quantity in children: the effect of vocabulary and participation. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 150–154.

Fraisse, P., and Piaget, J. *Traité de psychologie*, Vol. 7, pp. 101–102, Paris: Presses Universitaires de France, 1966.

Halford, G. S. An experimental test of Piaget's notions concerning the conservation of quantity in children. Journal of *Experimental Child Psychology*, 1968, **6**, 33–43.

Kendler, H. H., and Kendler, T. S. Vertical and horizontal processes in problem solving. *Psychological Review*, 1962, **69**, 1–16.

McLaughin, H. G. Psycho-logic: a possible alternative to Piaget's formulation. *British Journal Educational Psychology*, 1963, **33**, 61–67.

Piaget, J. *Logic and psychology*. New York: Basic Books, 1957.

Piaget, J. *The psychology of intelligence*. London: Reutledge and Kegan Paul, 1950.

Sutcliffe, J. P. A general method of analysis of frequency data for multiple classification designs. *Psychological Bulletin*, **54**, 134–137.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 328–355 (1969)

The Encoding and Decoding of Symbols by Children: A New Experimental Paradigm and a Neo-Piagetian Model[1]

JUAN PASCUAL-LEONE[2] AND JUNE SMITH
University of British Columbia

A semantic-pragmatic model for problem-solving and symbolic behavior is offered in which Piaget's cognitive-developmental variable is conceptualized as a quantitative construct, the central computing space M.

A new sort of concept attainment or decoding-encoding experimental paradigm based on the model is developed and an experiment on 5-, 7-, and 9-year-olds is reported. The experimental design includes three factors: chronological age, mode of representation (verbal vs. gestural), and semantic-pragmatic task complexity (decoding vs. encoding). The Piagetian predictions are upheld contradicting alternative predictions which follow from Bruner's model.

Finally, the new model is applied to explaining the Piagetian class-inclusion data and also the data of inferential behavior recently reported by Kendler and Kendler.

This paper considers two different but related problems: the pragmatic aspect of meaning and the relations between language skills and "intelligence" in Piaget's sense (or information-processing capacity or mediational power). These two problems will be discussed successively, and in the course of this discussion an experimental paradigm will be offered and specific predictions made. The second part of the first section presents a set of experiments which test the prediction. In the second and last section this neo-Piagetian model is used to explain the developmental data on Inclusion of Classes and the data on inferential behavior recently reported by Kendler and Kendler (1967).

Some Semantic-Pragmatic Concepts

Pragmatics is that little-developed part of semiotics which is concerned with the study of the relations between symbols (in the general

[1] The writing of this paper has been facilitated by grant NRC APA 234 (Canada) to the senior author. We are grateful to the Vancouver School Board and to the Vancouver schools Bayview and General Gordon for making the children available for these studies. The theoretical content of this paper is the sole responsibility of the senior author.

[2] Now at York University, Psychology Department, Toronto 12, Ontario (Canada).

sense) and their users. Pragmatics deals with problems such as the following: A person *produces* a sign; a person *receives* a sign; a person *understands* a sign *as it is understood* by the one who sent it; a person *acts* as a consequence of receiving a sign (Apostel, 1967). Technical concepts relevant for studying these problems are: symbol, decoding, encoding, belief, utterance, assertion, act, meaning (or "intension"), and object (or referent). Rigorous definitions for these concepts can be found scattered in places such as Tolman (1932, 1959), Carnap (1956), Apostel, Mays, Morf, and Piaget (1957), Piaget (1958), Osgood (e.g., 1963), Reitman (1965). Apostel (1967) has explicitly stated what seems to be the belief of the logicians and/or psychologists mentioned above, that the study of language is ultimately (but not exclusively) to be based in a semantic-pragmatic theory of (overt or covert) behavior. This view contrasts with the position adopted by two extreme groups by psycholinguists: those among the psychologist followers of Chomsky who would like to place syntax as the ultimate foundation of language and who so doing would place "a little linguist in every brain" (Osgood, 1963, p. 75[illegible] and those among the followers of Hull and Skinner who, by over-simplifying the psycholinguistic issues, ignore "all seven" of George A. Miller's admonitions (Miller, 1965, p. 18). Using this general approach as the frame of reference we turn now to the theoretical analysis of our experimental situations.

In the experimental paradigm presented below, two (or more) objects are displayed before the subject (*S*). The instructions differ according to the experimental condition. Under the "decoding" condition the experimenter (*E*) produces a symbol, verbal or gestural (e.g., says "red" or makes gesture of writing) and *S* has to infer to which one of the two objects the symbol refers. Under the "encoding" condition *S* has to produce a symbol (verbal or gestural) which will indicate to *E* which one of the displayed objects he has in mind. In either case after having detected or produced the symbol *S* will point to the object he considers to be referent. The analysis of these situations requires a definition of the concepts: "meaning," "object or referent," and "assertion."

Osgood defines meaning as "a bundle of simultaneous semantic features" (Osgood, 1963, p. 746). These semantic features or components have for Osgood the status of constructs and they seem to play in this regard a role similar to Piaget's "scheme" construct.[3] The notion of

[3] Piaget's French label *schème* is usually translated into English as "schema" and less frequently as "scheme." However, it is our opinion that the latter constitutes the only acceptable translation. The word "scheme" suggests a plan of action (internalized or external) which is precisely the core meaning of Piaget's predicative psycho-

scheme is similar to the Tolmanian notions of means-end readiness or hypothesis (Restle, 1967). (The scheme concept however is more general because it applies equally well to perceptual, emotional, or any other recurrent pattern of performance which is modificable through experience.) Piaget *et al.* (e.g., Apostel *et al.*, 1957, p. 50) explicitly define the meaning of a concrete object O_i for a subject S in a given situation as the union or the intersection of all the schemes in his repertoire which (from S's point of view) can successfully assimilate O_i. Piaget (1958; Apostel *et al.*, 1957) has carefully and explicitly defined in logical terms what he means by scheme and by assimilation. A scheme corresponds to a predicate function (i.e., a property or relation) which the S has stored in his repertoire and which is used by him for assimilating (i.e., interpreting or categorizing semantically or pragmatically) the input.

In sum, a scheme is in itself a token of the hypothetical construct "meaning" and its activation can be represented as a predicate expression such as $\phi(x)$ where x is an argument-variable standing for the scheme's referent. A scheme ϕ has in *principle* any degree of complexity, from a simple property such as "red" to a complicated property list which is hierarchically organized. According to Piaget's definition, any given meaningful object is a manifold category or a complex invariant defined by a cluster of schemes which are activated either at the same time or in a lawful invariant relation.[4] A category of objects can thus be represented as a compound-predicate expression such as $(\phi_1 \text{o } \phi_2 \text{o } \ldots \text{o } \phi_n\, x$, where the connective "o" stands for any sort of invariant relation. The meaning of any concrete object or referent (i.e., the psycho-logical class to which it belongs) is therefore a particularly complex scheme. When the input is assimilated (and hence an object O_i is "perceived") a "meaningful referent" Ω_i has been assigned to the input. From Piaget's point of view a compound-predicate function Ω_i existing in the S's repertoire has been satisfied by the features in the total input (feedback

---

functions or assimilatory schemes. The word "schema" in its modern usage suggests too strongly the idea of a template, a schematic form or an image. Following this distinction it is worth stressing with Piaget and Inhelder (1966, p. 431) that although schemas are nothing but figural-representational schemes, not all schemes are schemas. In fact, Piaget's scheme concept although more general, shows a surprising resemblance to the concept of operant as it is interpreted by some of Skinner's recent followers (see Staddon, 1967).

[4] The similarity between this psycho-logical construct of scheme and Carnap's concept of "intension" (Carnap, 1956, pp. 28, *ff.*) is worth noting. Those familiar with Cassirer's important contributions to the theory of knowledge will also recognize the striking similarity with Cassirer's notions of "concept" and of "object" (Cassirer, 1957, pp. 315, *ff.*).

included).[5] Thus, an object O_i as a cognitive entity can be defined by the expression:

$$\Omega_i(x), \text{ where } \Omega_i \equiv \phi_1 \circ \phi_2 \circ \ldots \circ \phi_n \quad (1)$$

An assertion is an *explicit* belief of a subject about a given state of affairs which is susceptible of being true or false. Carnap (1956, p. 249) suggests that "an assertion" is a construct inferable from the *S*'s behavior. In fact an assertion does not need to be communicated by a piece of verbal behavior. All sentences are evidently assertions but, as Osgood (1963) notes, not all assertions are sentences: Assertions can be perceptual (or representational) as well as verbal; "The sight of TOM DIPPING HIS HAND IN THE TILL (stealing) is just as much an assertion as the sentence *Tom is a thief*" (Osgood, 1963, p. 747).

In the context of the present situations: When *S*, after having heard the instructions, hears the utterance "red," notices that one of the perceived objects, for example, ϕ_1, satisfies the perceptual property red, and points to it, he has made an assertion equivalent to the sentence "The object that *E* has in mind is ϕ_1."[6]

It is trivially true that the activation of any scheme generates an assertion in the *S*. However it is not so trivial to observe that not all assertions are schemes available in the *S*'s repertoire. Just as the *S* can generate novel sentences, he can generate novel perceptual or cognitive assertions by composing schemes or cognitions within his repertoire. The processor or hypothetical construct (let us call it *M*) responsible for this productivity has been called an "Integrator Mechanism" (Maier, 1960), a finite "Central Computing Space" (Miller and Chomsky, 1963), a "Cognitive Mixer" (Osgood, 1963), a "Comprehension Operator" (Reitman, 1965), etc. Piaget calls it "Centration Mechanism" and/or "Field of Equilibrium" (Piaget, 1956). At any rate, the function of this central computing space is to "compose" or "mix" semantically (i.e., to coordinate or to integrate) a number of different schemes. It operates like a function of functions: $M(\phi_1, \phi_2, \phi_3, \ldots)$.

Two remarks are in order with respect to this central computing space or *M* operator: (a) Despite the widely held belief about verbal medi-

[5] Notice that not all the compound-predicate schemes (i.e., complex meanings) have a referent (i.e., an object). A well-formed sentence (or concept) has a meaning even if it is false (i.e., has no referent). This does not pose any problem for the present formulation because the nature of the input is left undefined. False concepts (or empty classes) may be functional invariants constructed by *S* in connection with the input constituted by free thought and/or by linguistic production.

[6] Following Carnap's (1956) similar procedure the symbol Ω_i is used here to represent both a compound-predicate function (or object-scheme) and its corresponding class (or any recognized member of it).

ation it does not make sense to say that M is verbal or linguistic: M belongs to a higher logical type than its linguistic product. In addition it operates equally well in nonlinguistic context. (b) Considerable evidence suggests that the power of M (maximum number of schemes included) increases with age. The hypothesis that the power value of M defines the different developmental stages of Piaget has in fact been put forward and successfully tested by the senior author (Pascual-Leone, 1968).

This introduces our second problem.

The Relations between Linguistic Skills and "Intelligence" or Information-Processing Capacity

Since we are identifying the construct "intelligence" with the size of the central computing space or M operator our position with regard to this issue should be clear already. This is, however, a controversial issue. There are many investigators in cognitive development (e.g., Kagan, 1966; Bruner, Olver, and Greenfield, 1966; etc.) who seem to imply that the *S*'s ability to think (i.e., *his ability to produce complex novel assertions of any sort*) is contingent upon his linguistic development. Inhelder and Piaget (see Inhelder and Piaget, 1964) and the present writers would suggest the contrary: that linguistic competence is a consequence of thought as defined above. Our decoding (D) and encoding (E) experimental situations were designed to test these two positions by means of the following derivations. Consider two types of task D and E such that the semantic complexity of the assertion required for succeeding D (i.e., the M operator "demand" of D) is, according to Piaget's theory, accessible to 5-year-olds, while that for E is accessible only to 7-year-olds or older children. Now consider (following Bruner *et al.*, 1966) two "systems" or "modes of representation," such as the "enactive" or gestural (G) mode and the "symbolic" or verbal (V) mode; the mode G is supposed to have been thoroughly mastered by 5-year-olds while the mode V is not mastered until 7 or 8 years. If *S*s from different age groups are tested under each of the four experimental conditions (i.e., DG, DV, EG, and EV) the Piagetian predictions should be that the variables "logical structure of the task" (i.e., D or E) and "chronological age" will interact independently of the "mode of representation"; furthermore, the Piagetian logical analysis should be able to predict the age of success for all four conditions. However, if Bruner's model is correct, the interaction will occur between "mode of representation" and "chronological age"; i.e., the logical analysis will be contradicted.

This paradigm seems to offer a critical test for the two positions. That is to say: If language is defined in its strict sense which is given

by any dictionary (i.e., the words, their pronounciation, and the methods for combining them used and understood by a considerable community), then language is different from logic and the paradigm proposed may help to decide between the Piagetian and the non-Piagetian interpretations. At the same time this paradigm offers a new minimally verbal context where Piaget's concept of "mental structures" can be tested.

The DV situation was described above as an illustration of a cognitive-perceptual assertion. A logical task analysis can give an estimate of the semantic complexity of this assertion: The subject must remember the instructions (call them ϕ_I); he must have explored visually the stimulus-display S constituted by objects O_1 and O_2 (call the corresponding perceptual-cognitive scheme ϕ_s); he must have heard the utterance $u\phi_i$ which stands as a name for the scheme (or concept) ϕ_i.

If S can coordinate all this information by means of his M operator, i.e., $M(\phi_I, \phi_s, u\phi_i)$, then the Cognitive Mixer of Osgood (1963) will operate generating the conclusion that ϕ_i belongs to the object-class or set Ω_1 which in its turn will lead to the S pointing to the corresponding concrete object O_1. A similar analysis can be done for the situation DG with the difference that now the "utterance" will be a gesture which must be identified by S with an action applicable to Ω_1. In summary, the two decoding conditions can be semantic-pragmatically defined by the assertion:

$$M(\phi_I, \phi_s, u\phi_i) \rightarrow p\Omega_1 \tag{2}$$

where the second term $p\Omega_1$ of the Cognitive-Mixer rewriting rule stands for the purpose of denoting object Ω_1 (i.e., the distal object O_1).

With regard to the encoding conditions a different functional structure appears. Now S is given the instructions I (which become ϕ_I after being decoded by S) and shown the stimulus-display S (i.e., ϕ_s from S's viewpoint); this time, however, the cue $u\phi_i$ is not available but *has to be discovered* (or constructed) by S using the objects Ω_1 and Ω_2. A cue $u\phi_i$ indicating unambiguously one of the objects must be such that the property ϕ_i designated by it is *included* in Ω_1, (i.e., $\phi_i \subset \Omega_i$) *but it is not* included in Ω_2 (i.e., $\phi \not\subset \Omega_2$).

In order to satisfy these two constraints any effective procedure will have to use the strategy of looking at the objects as if they were sets of specific properties (e.g., red, wooden, used-for-writing, etc.) which intersect (Figure 1). Granting that for the S any object is a compound-functional invariant of the type described in Formula 1, this strategy does not seem too far-fetched.

The S will have to classify these properties as belonging to the hatched area of Figure 1 (i.e., symmetric difference in set terminology) or to the

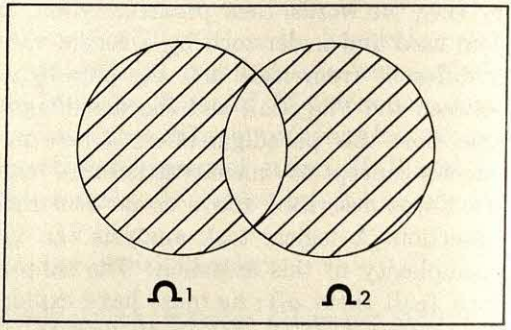


FIG. 1. Venn diagram of the relation between the property-sets which define the two objects Ω_1 and Ω_2 for the S.

nonhatched area in which Ω_1 and Ω_2 intersect. Only those specific properties which belong to one of the objects *but not* to both (i.e., symmetric difference) will satisfy the constraints.

In Piaget's terminology this classificatory activity is the logical Grouping VI of symmetrical relations (Flavell, 1963, p. 182). More precisely our example corresponds to the form of symmetrical relations which Piaget calls *altérités positives* (Piaget, 1949, p. 150, Def. 26). The interpretation of this grouping in the context of our experiment is better understood by referring to Figure 2. In this figure the Venn diagram does not represent sets of specific properties as in Figure 1 but classes (or sets) of objects. There is a general class B defined by a general property or relevant dimension Ψ_i such as (e.g.) color. This general class includes both a subordinate class A defined by a specific

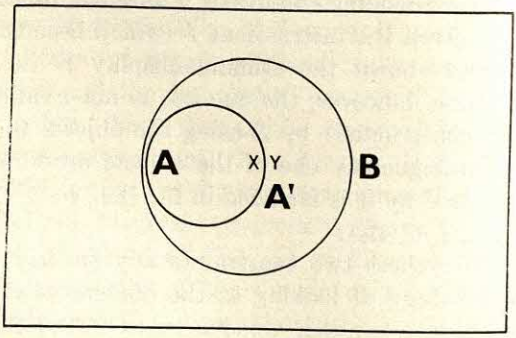


FIG. 2. Venn representation of a general class B of objects which includes two complementary classes: the class of objects A (x stands for one of these objects) and the class A' (i.e., the class of objects such as y which are B and *non-A*). The type of symmetric relation existing between objects x and y (i.e. sharing the same general property or dimension, but having a different specific property or value) is called by Piaget an *altérité positive*.

property ϕ_i such as red and a complementary class A' defined by the specific property ϕ'_i (e.g., color-other-than-red). The variables x and y of Figure 2 represent any two objects satisfying, respectively, the specific properties ϕ_i and ϕ'_i (i.e., x and y belong to the corresponding classes). These two objects x and y maintain with each other the symmetric relation defined by these two conditions: not sharing the same specific property and sharing the nearest general property (i.e., Ψ_i). In Piaget's language this symmetric relation is an *altérité positive*. It can be verified that our S's task can be reformulated à la Piaget in the following way: to find a pair of specific properties ϕ_i and ϕ'_i such that the objects Ω_1 and Ω_2 can satisfy (i.e., logically replace) the variables x and y in Figure 2. This is obviously one of the psychological jobs of Grouping VI. Incidentally, experts in Piagetian logic will have noticed that the diagram of Figure 2 can also be interpreted by means of Piaget's Logic of Classes as Grouping I of addition of classes. This is the better studied of Piaget's Groupings and constitutes the rationale for his well-known task called "Inclusion of Classes." This alternative interpretation of our task is important for the present predictions as will be noted below. Inhelder and Piaget (1959) have shown that this type of operations (i.e., M operations) is acquired during the concrete-operations period at 7–8 years of age and is well established at 9–10 years of age. If our analysis is correct the developmental comparison between decoding and encoding conditions offers a novel opportunity for testing Piaget's theory experimentally. This is our Hypothesis No. 1: *Independently of the type of cue used (verbal or gesture) the decoding conditions will be succeeded by 5- and 6-year-olds as well as by younger children (3- and 4-year-olds). However, the encoding condition will not be completely solved until 7–8 and not completely mastered until 9–10.*

A finer analysis of the encoding condition is possible by comparing the verbal vs. the gesture-cue responses. Under the gesture-encoding condition the child has to generate an appropriate gesture of an action which will suit Ω_1 but not Ω_2. Notice that for this purpose any physical feature of Ω_1 and Ω_2 is relevant because the actions applying upon any object are cued by the object as a whole. That is, the visual display constitutes a thoroughly "facilitating situation" in the sense that no irrelevant cue is provided (see Pascual-Leone and Bovet, 1966). Under these favorable circumstances the task can be solved without considering the general property Ψ_i. The findings of Inhelder and Piaget (1964) suggest that the ability to solve problems dealing exclusively with two different properties such as ϕ_i and ϕ'_i appears at 5–6 years. A more conclusive related result is offered by Kohnstamm (1963; 1967) who, using ingenious learning procedures, succeeded in transforming an Inclusion of

Classes task into a "facilitating situation" by way of extinguishing all the irrelevant misleading schemes. He found that under these conditions 5- and 6-year-olds can solve the task perfectly. Although Kohnstamm does not offer any logical interpretation of his data it appears from a semantic-pragmatic analysis that, under these facilitating conditions, the Inclusion of Classes is degraded into a task in which only two classes B and A (but not A') need to be considered. That is, again only two different predicate functions (i.e., properties) are used. The logical analysis of the class-inclusion task and of Kohnstamm's modification will be summarized in the last section of this paper. Now it will suffice to state that our encoding gesture condition can be assimilated, from a semantic-pragmatic point of view, to a facilitating-situation task of similar complexity. Here then is the Second Hypothesis:

With regard to the encoding procedure the gesture condition will be solved the first time at 5–6 years and thoroughly mastered at 9–10 years. Using the M operator construct the assertion corresponding to the decoding-gesture is:

$$M(\phi_I, \phi_s, \phi_i \subset \Omega_1, \phi'_i \subset \Omega_2) \rightarrow p\phi_i \tag{3}$$

where $p\phi_i$ stands for the purpose of producing the symbol corresponding to ϕ_i. It should be noted that the expression $\phi'_i \subset \Omega_2$ (i.e., "the specific property complementary of ϕ_i is included in the property-set Ω_2") is logically equivalent in our situation to the expression $\phi_i \subset \Omega_2$ (i.e., "ϕ_i is not included in Ω_2") which was used above for representing the second constraint.

A semantic-pragmatic analysis of the verbal encoding condition supports the previous conclusions. In order to find an appropriate verbal cue S has to find a perceptual cue in Ω_1 (or Ω_2) which can be labeled and which satisfies the two constraints $\phi_i \subset \Omega_1$ and $\phi'_i \not\subset \Omega_2$. But obviously there are many other perceptual cues in Ω_1 perhaps more salient which do not satisfy the constraints; these other perceptual cues become therefore distracting or even misleading. If the S is to have a chance to succeed he must have (while turning his attention from Ω_1 to Ω_2) a "searching image" (cf. Hinde, 1966, p. 89) or covert representational response for the general property he is "testing" (e.g., a representation for "color" when he is testing the hypothesis "red"). Thus, the assertion corresponding to the EV condition has to include the S's representation of the general property as a necessary and separate information (call the representation, Ψ_i):

$$M(\phi_I, \phi_s, \phi_i \subset \Omega_1, \phi'_i \subset \Omega_2, \Psi_i) \rightarrow p\phi_i \tag{4}$$

Notice that the complexity of content inside the parenthesis of the *M* operator increases steadily and regularly from Formula 2 to Formula 4. Elsewhere, the senior author (Pascual-Leone, 1968) has shown by means of a quantitative stochastic model that *M* formulas of the complexity of Formula 3 are accessible to the central computing space of 5- and 6-year-olds and *formulas of the type of Formula 4 are accessible to 7- and 8-year-olds but not to younger children. This is our Third Hypothesis.* It seems therefore that Piaget's model and the present semantic-pragmatic model agree with regard to these data.

METHOD

As mentioned above the experimental conditions were Decoding Verbal (DV), Decoding Gesture (DG), Encoding Verbal (EV), and Encoding Gesture (EG). For each one of these conditions the testing (T) period was preceeded by a training (t) period where *S* became acquainted with the task. All *S*s passed individually through all four conditions, according to either one of two randomly assigned orders of

TABLE 1

EXPERIMENTAL MATERIAL: DECODING VERBAL CONDITION: DV

Item	Objects	Cue	Matched dimension	Nonmatched dimensions
1.	Red block Red knife	Wood	Color	Size, material, use
2.	Plastic knife Steel knife	Silver	Use	Color, size, material
3.	Glass bowl Marble	Roll	Material	Size, use, color
4.	Needle Toothpick	Orange (color of toothpick)	Size	Color, material, use
5.	Hammer Ruler	Hit	Size, material	Use, color
6.	Plastic cup China cup	China	Use, size	Color, material
7.	Large plastic tumbler Small plastic tumbler	Big	Use, material	Color, size
8.	Pink plastic ice tray Pink plastic knife	Long	Color, material	Size, use

presentation: No. 1: tDV, TDV, tDG, TDG, tEV, TEV, tEG, TEG; No. 2: tDG, TDG, tDV, TDV, tEG, TEG, tEV, TEV.

Thus, the Decoding series always preceded the Encoding ones. The rationale behind this fixed order is to ensure that when the critical encoding conditions come the *Ss* have had full opportunity to: (a) demonstrate their linguistic competence by means of the decoding activity; (b) learn through imitation (or "vicarious learning") the sort of overt responses which will be required from them during the encoding period. Such a procedure, which obviously favors the null hypotheses, should rule out the interpretations of "lack of linguistic ability" or failure to "understand" the task which are often produced as objections to the Piagetian data.

An effort was made to select everyday objects familiar to any young child for each one of the four conditions. In each case a total of 12 different items were used, four during the (non-scorable) training period and eight during the testing period. A complete list of the items used for testing in each condition appears in Tables 1, 2, and 3. Items in the Gesture series were randomly paired while items in the Verbal series were intentionally matched in one (or two) specified dimensions and intentionally nonmatched in three (or two) specified dimensions. As in the present experiments only the nonmatched dimensions are relevant, this procedure introduces a redundancy factor in the verbal series which should facilitate the successful performance of EV condition. This design which favors again the null hypotheses would further guarantee the strength of our test should the experimental hypotheses be upheld. The four dimensions used for this purpose (i.e., size, color, material, and use) were counter-balanced with regard to their role as matched or nonmatched dimension. Table 1 presents also the cues used by the decoding conditions.

TABLE 2
EXPERIMENTAL MATERIAL: DECODING GESTURE CONDITION: DG

Item	Objects	Cue	Item	Objects	Cue
1.	Chocolate Eraser	Eat	5.	String Elastic band	Tie bow
2.	Shoe Sock	Step in shoe	6.	Egg cup Salt shaker	Sprinkle
3.	Brush Paper towel	Dry hands	7.	Face cloth Spectacles	Put on spectacles
4.	Needle Scissors	Cut	8.	Jar Match box	Unscrew lid

TABLE 3
EXPERIMENTAL MATERIAL: ENCODING CONDITIONS

Encoding Verbal: EV				Encoding Gesture: EG	
Item	Objects	Matched dimensions	Nonmatched dimensions	Item	Objects
1.	Sock Soap	Color	Size, use, material	1.	Apple Ball
2.	Plastic spoon Steel spoon	Use	Color, size, material	2.	Glove Hat
3.	Paper bag Writing paper	Material	Size, use, color	3.	Bracelet Lifesaver
4.	Pen Fork	Size	Color, use, material	4.	Comb Toothbrush
5.	Napkin Kleenex	Material, size	Color, use	5.	Orange Sugar cube
6.	Silver fork Plastic fork	Use, size	Color, material	6.	Candle Stick
7.	Long red pencil Short blue pencil	Use, material	Size, color	7.	Jug Flashlight
8.	Green plastic cup Green plastic fork	Color, material	Use, size	8.	Lollipop Barley sugar

Subjects[7]

Group 5 (i.e., the 5-year-old group) consisted of 24 kindergarten children, 15 of whom were boys. The mean age of the group was 5.11 years with *SD* of 3.5 months and a range of 5;6–6;6. Group 7 was formed by 24 second graders (13 boys) with a mean age of 7.11 years (*SD*, 3.3 months) and a range of 6;6–7;6. Finally, Group 9 constituted by 12 fourth graders (6 boys) having a mean age of 9.10 years, with a *SD* of 2.97 months and a range of 9;6–10;6.

Given the nature of the experimental hypotheses an effort was made to ensure that the children's intellectual performance was representative of their age. For that purpose a pool of 30 Group-5 children were pretested (in groups of four) with the Harris-Goodenough Figure Drawing Test (see Harris, 1963). The selected sample had a Harris-Goodenough mean IQ of 117 with a *SD* of 9.37 (range: 96–141). As Group 7 was the critical one from the point of view of Piaget's Theory the

[7] Prior to the experiments reported here a pilot study was conducted with ten 5-year-olds and ten 9-year-olds. The sample characteristics and the testing procedure were similar to the ones reported below. As these highly significant experimental results have been replicated by the larger study reported here, the pilot study will not be reported.

children were pretested with group versions of Piaget's Conservation of Horizontality or Water Level task (cf. Piaget and Inhelder, 1956, p. 375) and with an "Abelson Figures" (i.e., overlapping figures) test inspired by Piaget's "Intersection of Classes" task (Inhelder and Piaget, 1964, p. 176). Both group tests have been developed by the senior author in a series of unpublished studies. The reliability of the Water Level group test was for the present sample, $r = .67$ ($N = 24$; odd-even split-half Pearson correlation coefficient). The reliability of Abelson Figures administered as a group test was $r = .89$ ($N = 24$) for the present sample. Construct validity of the Water Level individual test has been demonstrated by the senior author in two unpublished factor-analytical studies interrelating different Piagetian tasks. The 24 children presenting the highest score of the total pool for both tasks were selected for study. The performance of these children on the Water Level task was congruent with the structural norms offered by Piaget and others as characteristic of their age groups. Children of Group 9 were selected by means of the Henmon-Nelson test of mental ability (Lamke and Nelson, 1957). The sample chosen showed an IQ average of 118.6 with an *SD* of 10.61.

All *S*s come from the same school in a middle-class neighborhood and belong to English-speaking homes.

Procedure

Decoding. The task was introduced as follows (training): "I am going to think of one of these two objects and I want you to guess which one I am thinking of. As a help I will give you a clue, this clue will be a word (or a movement). You can guess from my word (or my movement) which I am thinking of." An illustrative example followed which differed according to the condition. This is an example for the verbal-cue items: "Here are two balls of string. This one is tall and green. This one is short and white. If I say tall which one am I thinking of?" This other example corresponds to the gesture-cue items: "With this one I might pretend I was painting and make a movement like this (gesture); with this one I might pretend I was drawing a line and make a movement like this (gesture). If I do this (gesture) which am I thinking of?" For each one of the four training items the same procedure was followed. When the child failed to produce the right answer in any one of these four items new explanations and representations of the same item were offered until all four training items were passed.

During testing *E* merely presented the items and said: "If I say ——— (or "If I make a movement like this ———") which am I thinking of?"

Encoding. The training items were introduced as follows: "Now I

want you to think of one of these two objects and give me a word (or "make a movement") so that I can guess which you are thinking of." If *S* did not produce a cue the differential perceptual or gestural properties of the objects were drawn to the *S*'s attention and a training routine similar to the one outlined above was started. As before, *testing period was not introduced until S succeeded the four training items.* The name of the objects was not accepted as an adequate word-cue. The testing items followed introduced by *E*'s statement: "Think of one of these objects and give me a word (or "make a movement") so that I will know which you are thinking of." In order to further ensure that the *S*'s failures in the EV condition were not due to a lack of attention to the objects *E* pointed out to *S* his mistakes and he was encouraged to produce new responses to the item until he succeeded or until he had had three trials. On the first item *E* went as far as to provide *S* with a suitable cue should he be unable to form one for himself. However, only the first spontaneous response produced by *S* to each item has been considered in the scoring. The scoring procedure was the same for each one of the four experimental conditions; one point was given to the *S* for every one of the eight testing items successfully passed.

RESULTS AND DISCUSSION

As sex differences and differences in order of presentation were negligible these two factors have been disregarded in the analysis that follows.

Table 4 presents the main descriptive statistics: The mean score for each age-group and the corresponding *SD* together with the proportion of correct responses over the whole response pool. Figure 3 illustrates

TABLE 4

MEAN AND *SD* OF GROUP SCORES AND PROPORTION OF CORRECT RESPONSES

	Experimental conditions							
	DG		DV		EG		EV	
Group	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
5-year-olds	7.370	.807	7.540	.770	6.870	1.48	5.210	2.040
	(.922)		(.901)		(.859)		(.619)	
7-year-olds	7.650	.564	7.670	.552	7.210	.912	6.500	.957
	(.953)		(.958)		(.901)		(.812)	
9-year-olds	7.660	.471	7.580	.759	7.830	.553	7.500	.764
	(.958)		(.947)		(.979)		(.938)	

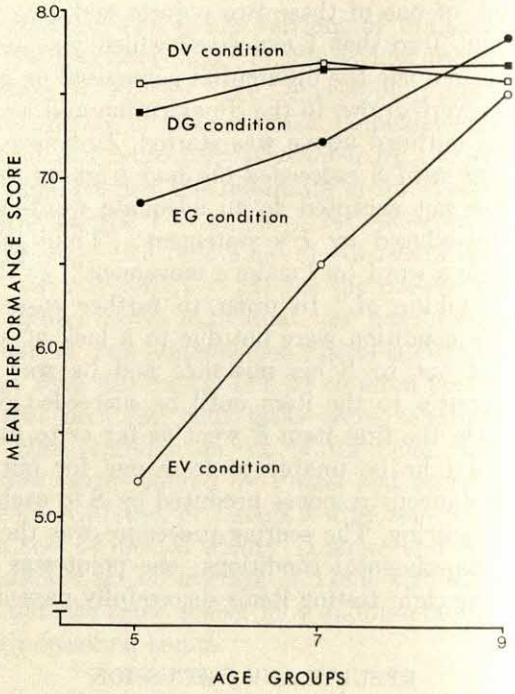


FIG. 3. Decoding (D) and encoding (E) abilities as a function of age and of "mode of representation": verbal (V) or gesture (G). For all conditions, the maximum score possible is 8: the score of 5.2 could be obtained (in EV) by chance.

graphically the mean performance of each age-group for the four conditions.

It appears that, in conformity with Hypothesis No. 1, the decoding conditions were perfectly mastered by all three groups; according to the Mann-Whitney U test, the differences of performance across groups are not statistically significant. With regard to the encoding series and for the condition EG the difference between Groups 5 and 7 is not significant (Mann-Whitney U test, $z = .699$) but there are highly significant differences between Group 5 and Group 9 (Mann-Whitney, $z = 2.776$, $p < .002$) and between Groups 7 and 9 (Mann-Whitney, $z = 2.557$, $p < .005$. These tests and the ones that follow are one tail). Even clearer differences are found for the EV condition for which the comparison of Group 5 vs. Group 7 shows significance of $p < .005$ (Mann-Whitney, $z = 2.567$) and of Group 5 vs. Group 9, which shows significance of $p < .0001$ (Mann-Whitney, $z = 3.642$). The comparison between Groups 7 and 9 is also highly significant (Mann-Whitney, $z = 2.761$,

$p < .003$). These results are in strong agreement with our Hypotheses Nos. 2 and 3 from which it follows that no difference between Groups 5 and 7 should be found for the EG condition but a significant difference should exist between these two groups for the EV condition.

A different set of predictions can be made from our hypotheses and, in general, from Piaget's concept of a concrete-operational stage of cognitive development. These predictions have to do with the within-group differences across conditions. They are two in number: (a) As according to Piaget and to our model the classificatory activity required by the tasks is thoroughly mastered by 9- and 10-year-olds, no difference should be found between decoding and encoding ability for this age group. (b) As 5- and 6-year-olds are expected to succeed the decoding and the gesture encoding conditions while the verbal encoding is only expected to be handled by 7- and 8-year-olds, a significant difference between EG and EV is predicted to occur in Group 5 but *no significant difference* is expected in Group 7.

The overall analysis of within-group differences was done by means of the Friedman two-way analysis of variance by ranks. As predicted no significant difference across the four experimental conditions was found for Group 9 ($\chi^2_r = 1.125$, df:3, *n.s.*) and highly significant differences were found for Group 7 ($\chi^2_r = 14.562$, df:3, $p < .01$) and for Group 5 ($\chi^2_r = 26.087$, df:3, $p < .001$).

Wilcoxon matched-pairs signed-ranks test was used to make the between-conditions comparisons for Groups 5 and 7. With regard to Group 5 no significant difference exists between DV, DG, and/or EG; however, a highly significant difference appears between any one of these conditions and EV (e.g., EG vs. EV, Wilcoxon, $T = 37.5$, $p < .005$). With respect to Group 7 no significant differences are found between DV, DG, and/or EG as well as between EG and EV; however, a highly significant difference exists between the decoding conditions and EV (Wilcoxon, $T = 20$, $p < .005$). The results are thus in agreement with predictions (a) and (b).

The reader will have observed that these main results have been directly analyzed in terms of specific effects without considering the main effects and interactions. This is because the semantic-pragmatic analysis was able to predict *directly* ordinal relations at the level of specific effects and therefore lends itself particularly well to a non-parametric analysis. However, it could be claimed that the general hypothesis stated above in the Introduction was formulated in terms of interactions. Indeed, the Piagetian models predict that the variables "logical structure of the task" (i.e., decoding or encoding; call it L) and "chronological age" (i.e., A) will interact independently of the

"mode of representation" (i.e., R). They also predict that the AR and the ARL interactions *will not* occur while the LR will. This is because *from a developmental viewpoint* and according to the Piagetian models the variable R is just an "apparent" variable which, as shown above, can be reinterpreted in terms of the "real" variable logical complexity.

Bruner's model, on the other hand, should predict that AR and ARL will interact as it considers the mode of representation to be a "real" developmental variable.

TABLE 5
ANALYSIS OF VARIANCE WITH REPEATED MEASURES

Source of Variation		*df*	*MS*	*F*
Between *Ss*				
SS_A	27.8067	2	13.9033	8.917*
$SS_{A/S}$	88.906	57	1.559	
Within *Ss*				
SS_L	45.938	1	45.938	51.042*
SS_{AL}	16.525	2	8.262	9.18*
$SS_{LS/A}$	51.323	57	.900	
SS_R	13.538	1	13.538	16.57*
SS_{AR}	2.634	2	1.317	1.611
$SS_{RS/A}$	46.616	57	.817	
SS_{LR}	18.706	1	18.706	15.87*
SS_{ALR}	7.054	2	3.527	2.141
$SS_{LRS/A}$	67.416	57	1.18	

* Statistical significance beyond $p < .001$.

Although the result of these interactions can be logically inferred from the analysis presented above it seems appropriate to test their existence directly. To that end an analysis of variance for three-factor experiments with repeated measures was conducted: L (decoding vs. encoding), A (5-, 7-, or 9-year-olds) and R (verbal vs. gesture). The corresponding results are summarized in Table 5. As can be observed, all the predictions laid out above are strongly supported by the data.

Some Secondary Supportive Results

Only one aspect not discussed deserves comment. This is the fairly high proportion of correct responses (i.e., .62, Table 4) produced by the 5-year-olds *vis-à-vis* the EV condition. As the model predicts that such a task will be too difficult for these children (see Hypothesis No. 3), the high proportion of successes remains to be explained lest it is taken as evidence against the model.

Three different empirical arguments can be advanced in favor of the

interpretation that this proportion of successes does not contradict Hypothesis No. 3. These three arguments are presented successively:

(1) An estimate can easily be made of the probability of obtaining by chance a correct response to the EV condition. It is given by computing in Table 3 (over the total of eight items) the proportion of nonmatched dimensions among the total number of matched or nonmatched dimensions. This proportion is .62, exactly the proportion of correct responses produced by 5-year-olds. Thus, according to a chance estimate, the correct-responding behavior of 5-year-olds is random.

This chance estimate is probably accurate because information obtained in a pilot study with regard to the relative salience of the different dimensions was used in selecting the item of the present study. The frequency with which a dimension appears as a nonmatched (or relevant) one is roughly proportional to its empirically found salience for 5-year-olds. Therefore the chance estimate is a conservative one.

(2) A different approach can be taken to demonstrate the noninferential character of the 5-year-olds' behavior. This is the study of the patterns of performance, or error-factors (see Tolman, 1959; Harlow, 1959; Restle, 1967) exhibited by the *S*s in their responses to the eight EV items. Most important in this regard is the existence of "response-sets" in connection with the type of dimension (or general property) employed by the *S* in his different answers (e.g., if *S* utters "red" his dimension is "color," if he utters "eating" it will be "use"). A measure of the response-set factor can be obtained by assigning to each *S* as a score value the number of utterances corresponding to the dimension he has used more frequently. This procedure generates a 7-point scale (i.e., from two to eight items).

The psychological interpretation of the response-set strategy is clear: By a simple recency effect the probability of responding to the same dimension should increase as a function of repetitions. That is, if no "mediational processes" or *M*-operated decentering activity occurs, the *S* will tend to fall into a response-set.

Table 3 shows that by merely using the response-set strategy a *S* could (in the case of color which is the most salient dimension) obtain as many as six correct answers. If our analysis is appropriate we should predict that a good number of correct responses obtained by 5-year-olds are generated by the nonmediated response-set strategy but that this will not be the case for older children. It follows that a significant positive correlation should be found between the number of correct responses and the response-set measure for 5-year-olds but no such correlation should exist for older *S*s. The Kendall Tau rank-order correlations computed on these data show for 5-year-olds a significant

result ($\tau = .292$, $N = 22$; $p < .03$);[8] however, the same correlations computed for Groups 7 and 9 are not significant ($\tau = -.09$ and $\tau = -.02$).

(3) If as suggested above the behavior of Group 5 is random with a bias introduced by the response-set tendency it should be expected that the distribution be unimodal and that it presents a very large variance. This follows because we are expecting to find a response-set bias applying over a Bernoulli distribution. Although mainly increasing the number of successes the bias could also increase the number of failures due to the response-sets in the matched dimensions. As Table 4 shows (EV condition) the *SD* of 5-year-olds is 2.04 while for 7- and 9-year-olds it is .957 and .764, respectively.

GENERAL THEORETICAL DISCUSSION AND CONCLUSIONS

The network of qualitative predictions drawn from the neo-Piagetian model has been upheld by the data.

Another compelling argument in favor of this model could have been made had we chosen to formulate the predictions quantitatively by means of probability theory. The *M*-operator model was conceived with this view in mind and the senior author has shown elsewhere (Pascual-Leone, 1968) how these quantitative derivations can be made. Instead we shall try to show now (remaining within the limits of semantic-pragmatic analysis) the generality of our model. This does not need to be done in the domain of psycholinguistics since the work of Miller and Chomsky (1963) among others has already shown that the notion of a central computing space *M*, equivalent to our construct, is necessary for understanding the linguistic performance of a language user.

It would be more relevant to outline how the present approach can explain piagetian data (e.g., the class inclusion task) or some recent behavioristic developmental findings such as the ones reported by Kendler and Kendler (1967). These two issues will be discussed successively.

The Inclusion of Classes Task

The class inclusion problem can be operationalized in different ways. The most difficult form and also the better studied is called by Piaget "Quantification of Inclusion." (For a summary review of some of the literature see Kohnstamm, 1967; also Wohlwill, 1968.) The logical

[8] Two children were eliminated because, despite the instruction and the training procedure, they answered every EV item with either the name of an object or with a gesture. The name and/or a gesture are neither correct responses nor acceptable dimensions.

structure of this task can be described by means of Figure 2. In this figure the variables x and y refer to two "individuals" (say a bird and a dog) which belong respectively to the classes A (e.g., Birds) and A' (e.g., Animals-non-birds). Obviously the union of A and A' is the general class Animals (i.e., B). The (quantification of) class inclusion task consists in presenting to S a number of instances of X, say $x_1, x_2, \ldots x_n$, together with instances of Y (i.e., $y_1, y_2, \ldots y_m$, where $n \geqslant m$) and then asking the standard question: "Are there more B or more A?" (e.g., "Are there more animals or more birds?"). The two sets of items X and Y can be presented pictorially or verbally; their number can be the same (i.e., $n = m$) or different (i.e., $n > m$); their composition can be homogeneous (e.g., X = four sparrows; Y = four dogs) or heterogeneous (e.g., X = four birds of different types; Y = four different animals, non-birds); some other set Z of extraneous elements $z_1, z_2, \ldots$, can be added to the display (e.g., a telephone, a tree, . . . together with sets X and Y). All these variations, although in a reliable way changing the performance level of Ss, alter neither the basic developmental difficulty of the task nor the error patterns in the S's responses.

An examination of these error patterns shows that those children (and some adults!) who fail the task, typically decode the question "Are there more B or more A?" as meaning "Are there more A or more A'?" Thus, when the perceptual display shows that the set X is larger than set Y (i.e., $n > m$), Ss will assert that "there are more A" and justify their assertion by referring to m and n. Similarly in experiments where the number of elements in X has been made equal to the corresponding number in Y (i.e., $n = m$), the Ss assert the equality of A and B and justify as above. Pascual-Leone and Bovet (1966) have pointed out that this behavioral pattern suggests the existence in the Ss' repertoire of a misleading semantic scheme (i.e., a cognitive error-factor or semantic habit). This semantic scheme was defined by them as a bias towards interpreting the two nouns (or arguments) mentioned in the question (i.e., B and A) as referring to the two distinct collections of objects available in the visual display (i.e., A' and A). The semantic bias was supposed to be facilitated by two factors: (1) the fact that one of the elements (i.e., A) is common to both the visual and the linguistic input; and (2) the semantic habit acquired during everyday social interactions where questions are organized in such a way that the correspondence between the components of the visual display and the nouns of the question is assured. This explanation appears to be acceptable to Kohnstamm in his recent monograph (1967, p. 26), and to Wohlwill (1968) who presents a similar construct under the name

"faulty perceptual set." Evidence in support of such a construct can be obtained by testing predictions about the generalization and the strength of this habit as a function of the number of releasing cues available in the situation. Data clearly interpretable from this point of view has been provided by Smedslund (1964), Pascual-Leone and Bovet (1967), Kohnstamm (1967), and, particularly, Wohlwill (1968). Call this misleading semantic scheme λ_e and consider next how it is possible for S ever to solve the class-inclusion problem if λ_e is such an enduring behavioral disposition in his repertoire. In order to facilitate the discussion it is convenient to introduce the following functional notation: call ϕ_a, $\phi_{a'}$, and ϕ_b the S's perceptual visual schemes (or *schemas*) corresponding respectively to the concepts or logical classes A, A', and B; call λ_a, $\lambda_{a'}$, and λ_b the listener-speaker's linguistic schemes corresponding to the same three concepts; call Ω_a, $\Omega_{a'}$, and Ω_b the manifold categories or superordinate conceptual schemes which in the S's repertoire stand for the logical classes A, A', and B themselves. Finally, call ϕ_I the S's representation of the task instructions given by E and call ϕ_s the S's representation of the visual display.

By using these notational definitions and the model of the class-inclusion task described above the following true expressions ("rules of truth," Carnap, 1956) can be derived:

$$\lambda_e \equiv (\phi_I, \phi_s) \supset \{(\lambda_b \overset{\sigma}{\rightarrow} \Omega_{a'}) \wedge \neg (\lambda_b \overset{\sigma}{\rightarrow} \Omega_a)\} \tag{5}$$

$$\Omega_b \equiv (\Omega_a \vee \Omega_{a'}) \tag{6}$$

$$\Omega_a \equiv (\ldots \circ \lambda_a \overset{\sigma}{\equiv} \phi_a \circ \ldots) \tag{7}$$

$$\Omega_{a'} \equiv (\ldots \circ \lambda_{a'} \overset{\sigma}{\equiv} \phi_{a'} \circ \ldots) \tag{8}$$

$$\Omega_b \equiv (\ldots \circ \lambda_b \overset{\sigma}{\equiv} \phi_b \circ \ldots) \tag{9}$$

$$\lambda_b \equiv (\lambda_a \vee \lambda_{a'}) \tag{10}$$

$$\phi_I \equiv (\ldots \circ \lambda_a \circ \lambda_b \circ \ldots) \tag{11}$$

$$\phi_s \equiv (\ldots \circ \phi_a \circ \phi_{a'} \circ \ldots) \tag{12}$$

Where the meaning of the connectives, $\neg$ (negation), $\equiv$ (equivalence), $\wedge$ (conjunction or logical product), $\vee$ (disjunction or logical addition), $\supset$ (implication), is the usual one in elementary functional logic. The meaning of the "semantic equivalence" $\overset{\sigma}{\equiv}$ is self-evident; the "semantic function" $\overset{\sigma}{\rightarrow}$ indicates that its first term (or antecedent) has its second term as a semantic referent; finally, $\circ$ stands, as in Formula 1, for any sort of invariant relation.

Before proceeding further it may be important to outline the inter-

pretation of Formulas 5–12. Formula 5 simply restates in symbolic language the description of λ_e given above including its facilitating factor No. 1. Formula 6 restates in the language adopted here the well-known Formula of Piaget's for Grouping I of addition of classes: $B = A + A'$ (see Figure 2). Formulas 7–9 define the conceptual schemes Ω_i by means of compound-predicate expressions in which all the constituents but two have been represented by suspension points. The formulas state that both the linguistic scheme λ_i and the perceptual scheme ϕ_i are included in Ω_i (they function as cues for Ω_i) and that from a semantic point of view they are equivalent. Formula 10 does not require any comment except to note that it follows as a theorem from Formulas 6, 7, 8, and 9. Formulas 11 and 12 signify by means of the conventions used above the S's representation of the task instructions (e.g., ϕ_I) and his perception of the visual display (i.e., ϕ_s).

Any reader keeping his cool despite the mathematical symbolism will have recognized that these semantic rules are little more than a restatement of the functional structure of the class-inclusion task described above. The distinct advantage of this symbolic formulation is that it can be considered a description of the task *from the S's point of view,* when the symbols are interpreted in the light of the neo-Piagetian model discussed earlier in this paper. That is: It can be assumed on the basis of considerable empirical evidence (see Wohlwill, 1968; Kohnstamm, 1967) that *S*s as young as 5½ years old already possess in their repertoires cognitive schemes corresponding to each one of the terms defined in Formulas 5–12. This assumption entitles us to find out by logical derivations whether it is possible to generate the correct answer ($A < B$) to the inclusion task (i.e., "there are more B than A because all A are B") in spite of the misleading influence of λ_e. According to the neo-Piagetian model such a production of this novel correct response (i.e., $R_{A<B}$) would demand an "integration" or "cognitive mixing" of the schemes the S has available in his repertoire. A more explicit definition of the "cognitive mixing" rule is thus required. For the present purposes consider the following five postulates as a definition of the "cognitive mixing" rule: P1: The overt (perceptual or behavioral) response R produced by an S in any given situation S^* is a complex function of all the schemes activated in S at the moment of responding. P2: All compatible schemes activated in S^* will apply together and co-determine R. P3: Other things being equal, when two or more totally incompatible schemes are activated in S^*, only the one compatible with the greatest number of other activated schemes will codetermine R. The other incompatible schemes will be inhibited. P4: Those schemes which are *partly* incompatible with the larger set of compatible activated

schemes will still apply by "accommodation," i.e., by inhibiting those among their subordinate alternatives which are incompatible and by activating the best compatible alternative available, if any. P5: Two or more schemes are said to be compatible when the logical expressions representing each of them appear to be logically consistent.

In the light of modern learning theories the five postulates do not appear to be revolutionary and they have the virtue of being congruent with the Piagetian theorizing (see Apostel, 1959, p. 67; Piaget, 1959, p. 47).

Now using these postulates and the semantic Formulas 5 to 12 the reader can easily verify that:

$$\{\lambda_e, \phi_I, \phi_s, \Omega_a, \Omega_b, \Omega_{a'}\} \rightarrow T_{A<B} \tag{13}$$

that is, the set of *activated* schemes on the left side of Formula 13 will, when logically combined, generate the cognitive map T (Tolman, 1948; Restle, 1967) leading to the correct answer to the inclusion task ($A < B$). *However, if the scheme $\Omega_{a'}$ had been excluded from this set, the incorrect answer $A \geqslant B$ would have been generated.* This is because, as Formulas 6 and 10 suggest, the label λ_b can be interpreted either as applying to both Ω_a and $\Omega_{a'}$ or as applying to $\Omega_{a'}$ only. (The third possible alternative does not arise in the context of the instructions ϕ_I because a reference to Ω_a, i.e., λ_a, is already being made.) Thus, when $\Omega_{a'}$ *is not* activated the incompatibility between λ_e and Ω_b can be solved (according to P4) by interpreting λ_b as referring exclusively to $\Omega_{a'}$. Such a compromise is not possible when the S has also activated the scheme $\Omega_{a'}$ because this scheme does not include the label λ_b (see Formula 8). According to P3 λ_e should therefore be inhibited and the answer $A < B$ produced. Notice that of all the schemes included in the first term of Formula 13 *only λ_e is an overlearned habit with a high probability of activation*. The other schemes will require from the S an act of selective attention in order to be *simultaneously* activated. "Selective attention" means in the neo-Piagetian model the use of the central computing space M. The realization by the S of the solution described by Formula 13 will therefore demand of him to perform the following M operation:

$$M(\phi_i, \phi_s, \Omega_a, \Omega_b, \Omega_{a'}) \rightarrow pR_{A<B} \tag{14}$$

where $pR_{A<B}$ stands for the purpose of uttering the correct answer. Two important observations should be made with regard to Formula 14. One is its congruence with Piaget's theory, according to which the Inclusion of Classes is solved by intermediate of the mental structure called Grouping I (addition of classes). As mentioned above Formula 6 is

nothing but this Grouping I written in a different manner. The second observation to be made about Formula 14 concerns its comparison with Formula 4. Notational and content differences notwithstanding the two formulas are isomorphic in the sense that their respective elements can be put in one-to-one correspondence. Indeed both formulas have the structure of M operations and both exhibit within the parentheses the same number of constituents: ϕ_I, ϕ_s, and three other "mediated" schemes or chunks of information. As mentioned above the M-operator model predicts that this complexity of information processing is only attained at about 7 to 8 years of (mental) age. The abundant literature reviewed by Kohnstamm (1967) shows indeed that this is the case for the present task.

Next we shall proceed to prove that the teaching of "didactic learning" procedures used by Kohnstamm (1963, 1967) and others can change the semantic-pragmatic structure of the class inclusion problem making it accessible to 5- and 6-year-olds.

Kohnstamm's didactic learning consists in exposing the child to a great number of class-inclusion tasks presenting the logical structure described above but showing different concrete content. In all tasks, E used explicit verbal reinforcement ("no, that's not right"; "yes, that's right") followed and preceeded by recurrent concrete verbal explanations which can be epitomized by means of Kohnstamm's own words (1967, p. 55): "You have to say that there are more B because A are also B. A and A' are *all* B and so there are always more B." It is thus fair to say that the teaching was specifically aimed towards "extinguishing" the misleading semantic scheme λ_e (Formula 5) and creating instead the facilitating semantic-pragmatic scheme λ_k:

$$\lambda_k \equiv (\phi_I, \phi_s) \supset \{(\lambda_b \overset{\sigma}{\rightarrow} \Omega_a \vee \Omega_{a'}) \wedge (pR_{A<B})\} \tag{15}$$

That is, a learning-set such that the decoding of the class-inclusion situation (i.e. ϕ_I, ϕ_s) will cue the S towards: (1) interpreting the *label* B (i.e. λ_b) as referring to *both* A and A', and (2) replying to the class-inclusion question by using the correct answer.

Now consider those 5- or 6-year-olds who successfully underwent this didactic learning. It is obvious that the internal state of the black box (i.e., the mind) of these children has changed. Before the teaching the children's repertoire pertinent to the task (i.e., relevant or misleading schemes) has been *summarily* represented by Formulas 5 to 12. However, after the teaching Formula 5 has to be eliminated and replaced by Formula 15. With the new repertoire, Formula 13 can no longer be derived and Formula 14 although still possible becomes unparsimonious. Indeed within a "total field of activation," i.e. { . . . }, including

scheme λ_k and excluding λ_e the correct response can be generated as follows:

$$\{\lambda_k, M(\phi_I, \phi_s, \Omega_a, \Omega_b)\} \rightarrow pR_{A<B} \tag{16}$$

This assertion can be easily verified by consulting the formulas of the new repertoire. A comparison with Formula 3 (from which the "total field of activation" has been excluded) shows the same number of schemes being integrated by the M operations. The M operations are isomorphic: both should be accessible to 5- or 6-year-olds. This explains Kohnstamm's data and concludes the proof.

The interesting series of control experiments conducted by Kohnstamm (1967) can be cited, independently of the author's interpretation, in support of Formulas 15 and 16.

The Inferential-Behavior Task of Kendler and Kendler

The explanatory power of the M-operator model can perhaps be better assessed when tested against unexplained new behavioristic data on the development of thinking. The best example is probably the brilliant series of experiments carried out by Kendler and Kendler (1967). The paradigm used in these experiments is a two-step one. In the first step children learn three simple schemes (or S-R associations): (1) To push a button A in order to be rewarded with a marble, B. (2) To push a button X in order to get a reward Y. (3) To put a marble B' (similar to B) handed by the E, into a slot in order to obtain the goal-reward G.

During the second step the child is faced with the simultaneous presentation of the panels corresponding to 1., 2., and 3., and is asked to obtain G *but is not given any marble* B'.

The logical interpretation of the task appears straightforward due to the refined experimentation of Kendler and Kendler. In order to achieve the task the S must be capable of self-generating a logical inference or assertion of the form:

$$\{A \supset B, B' \supset G, B \equiv B'\} \rightarrow A \supset G$$

that is: If A implies B (Scheme No. 1) and B' implies G (Scheme No. 2) and B is equivalent to B', then A necessarily implies G. This is a tautology in the Logic of Propositions. Looking at the task with piagetian eyes this tautology becomes a model for what is going on (consciously, or unconsciously) inside those Ss who succeed the task. Of the three premises of the tautology the two implications have been experimentally built into the S during step one. The equivalence $B \equiv B'$ must have been spontaneously (and covertly) built by S. Convincing experimental

evidence in support of this interpretation is offered by Kendler and Kendler's careful experimental analysis. In fact our logical equivalence scheme can be considered an "explication" (in Carnap's sense, 1956, p. 7) of Kendler and Kendler's own construct, the hypothetical representational response common to both segments (i.e., to Scheme Nos. 1 and 2).

Now it is possible to write down the formula for the logical-but-not-verbal assertion required from the subjects if they are to succeed the task:

$$M(\phi_I, \phi_s, \phi_{AB}, \phi_{B'G}, \phi_{BB'}) \rightarrow pR_{AG} \qquad (17)$$

The notation is the same used above in Formulas 2, 3, and 4. Five independent "chunks" of information (i.e., separate schemes) are integrated by the S's M operator: The instructions (ϕ_I), the stimulus display (ϕ_s), the two implications and the equivalence. The necessary psychological consequence of this integration is the intention of pushing-button-A-in-order-to-get-G.

The reader will have noticed that the Formula 17 is isomorphic with Formulas 4 and 14 discussed above. It was mentioned there that according to Piaget's theory and/or to our model an M formula of this complexity is only accessible to the central computing space of 7- and 8-year-olds but not to that of younger children.

Thus the Piagetian models predict the results obtained by Kendler and Kendler which according to their explicit admission *are not* explained by the neo-Hullian approach.

A word should be said about the possible connections between the data and the model presented above, and the role-taking behavior studied developmentally by Flavell and others (see Flavell, 1968). The need for this remark arises from the fact that our decoding-encoding paradigm could be interpreted as a role-taking situation.

Such an interpretation obviously suggests the possibility of using the M-operator model for explaining the development of role-taking behavior. The explanation could perhaps become the "causal-analytical" or "antecedent-consequent" theory which Flavell and associates (1968) consider a necessary but as-yet-to-come complement of the "developmental-descriptive" theory they present in their book.

This paper and its discussion of experimental results were presented mainly to serve two related purposes. The first purpose was to emphasize the possibility and the interest of an experimental-and-theoretical approach to language and thinking which is psycho-logical and semantic-pragmatic in its method of model building. According to this approach thinking, as well as the structural linguistic abilities studied by generative-transformational grammars, should not be confounded with

the verbal skills. Thinking and the structural linguistic abilities are considered here to have a logical (*but not necessarily verbal*) developmental and functional core.[9]

The second purpose of the present paper was to show that Piaget's developmental theory of intelligence when "liberalized" and interpreted in functional semantic-pragmatic terms, is able to account for data that neither the non-piagetian cognitive approaches nor the neo-behaviorist theories can nowadays explain.

REFERENCES

Apostel, L. Syntaxe. Sémantique et pragmatique. In J. Piaget (Ed.), *Logique et connaissance scientifique*. Paris: Gallimard, 1967. Pp. 290–311.

Apostel, L., Mays, W., Morf, A., and Piaget, J. *Les liaisons analytiques et synthétiques dans les comportments du sujet*. Paris: Presses Universitaires de France, 1957.

Apostel, L. Logique et apprentissage. In L. Apostel, A. R. Jonckheere, and B. Matalon (Eds.), *Logique, apprentissage et probabilité*. Paris: P. U. F., 1959. Pp. 1–138.

Bruner, J. S., Olver, R. R., and Greenfield, P. *Studies in Cognitive Growth*. New York: Wiley, 1966.

Carnap, R. *Meaning and necessity*. Chicago: University of Chicago Press, 1956.

Cassirer, E. *The philosophy of symbolic forms*. Vol. 3. *The phenomenology of knowledge*. New Haven: Yale University Press, 1957.

Flavell, J. H. *The developmental psychology of Jean Piaget*. New York: Van Nostrand, 1963.

Flavell, J. H. *The development of role-taking and communication skills in children*. New York: Wiley, 1968.

Harlow, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959. Pp. 492–537.

Hinde, R. A. *Animal behavior*. New York: McGraw-Hill, 1966.

Inhelder, B., and Piaget, J. *The early growth of logic in the child*. New York: Harper & Row, 1964.

Kagan, J. A developmental approach to cognitive growth. In H. J. Klausmeier and C. W. Harris (Eds.), *Analyses of Concept Learning*. New York: Academic Press, 1966. Pp. 97–115.

Kendler, T. S., and Kendler, H. H. Experimental analysis of inferential behavior in children. In L. P. Lipsett and C. C. Spiker (Eds.), *Advances in child development*. Vol. 3. New York: Academic Press, 1967. Pp. 157–190.

Kohnstamm, G. A. An evaluation of part of Piaget's theory. *Acta Psychologica*, 1963, **21**, 313–356.

Kohnstamm, G. A. *Teaching children to solve a Piagetian problem of class inclusion*. Amsterdam: North-Holland, 1967.

Lamke, T. A., and Nelson, M. J. *Examiner's manual. The Henmon-Nelson tests of mental ability*. Boston: Houghton, 1957.

[9] Some pilot experiments directly relating the structural linguistic abilities with Piaget's theory of intelligence have already been done by the Geneva School. This is the purpose of a doctoral thesis (H. Sinclair-de Zwart, 1967) and of a book in preparation by B. Inhelder and H. Sinclair-de Zwart (personal communication).

MAIER, N. R. F. Selector-integrator mechanisms in behavior. In J. Kaplan and S. Wapner (Eds.), *Perspectives in psychological theory*. New York: Inter. Univ. Press, 1960. Pp. 135–172.

MILLER, G. A. Some preliminaries to psycholinguistics. *American Psychologist*, 1965, **20**, 15–20.

MILLER, G. A., AND CHOMSKY, N. Finitary models of language users. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 2. New York: Wiley, 1963. Pp. 419–491.

OSGOOD, C. E. On understanding and creating sentences. *American Psychologist*, 1963, **18**, 735–751.

PASCUAL-LEONE, J. A mathematical model for the transition rule in Piaget's developmental stages. Research Report, Grant NRC APA 234 (Canada), 1968. York University. *Mimeo.*

PASCUAL-LEONE, J., AND BOVET, M. C. L'apprentissage de la quantification de l'inclusion et la théorie opératoire. *Acta Psychologica*, 1966, **25**, 334–356.

PASCUAL-LEONE, J., AND BOVET, M. C. L'apprentissage de la quantification de l'inclusion et la théorie opératoire. Partie II. *Acta Psychologica*, 1967, **26**, 64–74.

PIAGET, J. *Traité de logique*. Paris: Colin, 1949.

PIAGET, J. Logique et équilibre dans les comportements du sujet. In L. Apostel, B. Mandelbrot, and J. Piaget (Eds.), *Logique et équilibre*. Paris: Presses Universitaires de France, 1957. Pp. 27–117.

PIAGET, J. Assimilation et connaissance. In A. Jonckheere, B. Mandelbrot, and J. Piaget (Eds.), *La lecture de l'expérience*. Paris: Presses Universitaires de France, 1958. Pp. 49–108.

PIAGET, J. Apprentissage et connaissance. In P. Grèco and J. Piaget (Eds.), *Apprentissage et connaissance*. Paris: P. U. F., 1959. Pp. 21–67.

PIAGET, J., AND INHELDER, B. *The child's conception of space*. London: Routledge, 1956.

PIAGET, J., AND INHELDER, B. *L'image mentale chez l'enfant*. Paris: Presses Universitaries de France, 1966.

REITMAN, W. R. *Cognition and thought: An information processing approach*. New York: Wiley, 1965.

RESTLE, F. Linear theory of performance. *Psychological Review*, 1967, **74**, 63–70.

SINCLAIR-DE ZWART, H. *Acquisition du langage et developpement de la pensée*. Paris: Dunod, 1967.

SMEDSLUND, J. Concrete reasoning: A study of intellectual development. *Monographs of the Society for Research in Child Development*, 1964, **29**, 1–39.

STADDON, J. E. R. Asymptotic behavior: the concept of operant. *Psychological Review*, 1967, **74**, 377–391.

TOLMAN, E. C. *Purposive behavior in animals and men*. New York: Century, 1932.

TOLMAN, E. C. Cognitive maps in animals and men. *Psychological Review*, 1948, **55**, 189–208.

TOLMAN, E. C. Principles of purposive behavior. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 1. New York: McGraw-Hill, 1959. Pp. 92–157.

WOHLWILL, J. F. Responses to class-inclusion questions for verbally and pictorially presented items. *Child Development*, 1968, **39**, 449–465.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 356–365 (1969)

Concurrent Schedules of Social Reinforcement and Dependency Behavior among Four-Year-Old Children[1]

DAVID C. SPEER
Amherst H. Wilder Foundation

PETER F. BRIGGS
University of Minnesota

AND

ROCHELLE GAVALAS
Brain Research Institute, University of California, Los Angeles

An experiment exploring the validity of the Gavalas and Briggs concurrent schedules of social reinforcement model of dependency is reported. In a puzzle completion task, two responses were defined as follows: *dependent* (R_D), *S* obtaining a puzzle piece from the *E* who was seated some distance away; and, *competent* (R_C), *S* obtaining a puzzle piece by himself. Experimental groups received different concurrent schedules of social reinforcement for R_C and R_D. Analysis of variance yielded few reliable experimental group differences. Girls emitted reliably more R_D during extinction than did boys, independent of schedule of reinforcement. However, correlational analysis indicated that the model predicted the order effects of the concurrent schedules during extenction on a measure of R_D relative to R_C for sexes combined, but more reliably for girls than boys. The model also reliably predicted order effects on a mean rate of R_C for both sexes during extinction, but the relationship was stronger for boys than girls.

Social learning theorists have tended to group into one response class, dependent behaviors with differing response topographies. They also agree that dependent behavior is supported by other people who nurture, reassure, help, attend, satisfy or comfort the individual behaving in a "dependent" fashion. The reader is referred to Bandura and Walters (1963), Hartup (1963), and Walters and Parks (1964) for discussions of dependency exemplifying these views. The inclusiveness of these characteristics of dependency is insufficient for a tight definition. This looseness led

[1] This study was supported in part by Social and Rehabilitation Service Grant No. RT-2.

Gavalas and Briggs (1966) to suggest a new social learning model of dependency. Briefly, Gavalas and Briggs define dependency as follows: "Dependent behavior is persistently recurring behavior characterized by the seeking of others and by a lack of achievement or success, which has been caused by an earlier history of reinforcement for seeking others, concurrent with a lack of reinforcement for competency." (Gavalas and Briggs, 1966, p. 101) *Seeking of others* suggests that an operational behavioral measure of approach to another person would serve as one leg of a behavioral criterion for dependency. *Lack of achievement* suggests that a measure of relative effort such as distance traversed or time consumed would serve as another operational characteristic of dependency.

The concepts of achievement or success imply, however, a relatively specific situation context with some type of task connotation and associated S^D's or CS's. Thus, something about the situation should provide a stimulus to signal imminent reward for inefficient other-seeking behavior. For example, in a nursery school situation, when an adult announces a free play period, this may mean to some children "you are to seek or obtain pleasure in any way you wish, but you are to take the initiative in doing this." Many children will seize this as an opportunity to explore, exert mastery, or create. The "dependent" child, however, is likely to remain passively inert, and thus obtain encouragement and invitations to play from the adult. The dependent child may come to the adult with a multiplicity of requests for assistance and "I don't know what to do's," thus obtaining adult attention, conversation, or assistance. The change in the antecedent stimulus context may provide the discriminative stimuli previously associated with dependency reward and competency nonreward in the experiences of these latter children.

The fourth element in this model of dependency is competition between the competent and dependent responses generated by concurrent schedules of reinforcement. The distinguishing feature of a concurrent schedule is that two (or more) mutually exclusive operant behaviors are subject to two (or more) independent schedules of reinforcement. The organism is free to emit both operants but only in sequence. Reinforcement dispensed on schedule A is contingent upon the emission of operant A (and not B) while reinforcement dispensed on schedule B is contingent on the emission of operant B (and not A) (Ferster and Skinner, 1957; Catania, 1966). Gavalas and Briggs suggest that training and maintenance of dependent behavior under a concurrent schedule occurs where social reinforcement is dispensed at a high rate for inefficient, other-seeking behavior *and* at a low rate for efficient, autonomous, competent behavior. These authors suggest that the absence of reinforcement for competence is probably a more potent determiner of dependent responding than is the reward for

dependent behavior itself. Support for this view is found in the work of Catania (1966) in which animals were used as subjects. Catania and others have demonstrated that although the rates of the competing operants are relatively independent of one another there is an inverse relationship between the rate of emission of one operant and the rate of *reinforcement* of the competing operant. Thus, by holding the schedule of reinforcement for operant A constant, the frequency of operant A will vary inversely with the schedule of reinforcement for operant B.

The present paper reports an attempt to experimentally explore the adequacy of the Gavalas and Briggs model in describing and explaining the dependent behavior of preschool children. The approach involves stipulating different concurrent schedules of social reinforcement for competent responses (R_C) and dependent responses (R_D). Each schedule was imposed upon a group of subjects.

METHOD

Apparatus

Twenty colorful puzzles with the main object cut out (e.g., a chair, wagon, Christmas tree, elephant, etc.) were prepared of canvas board glued to masonite. On the masonite were lines to form 1 × 1-inch squares to show the location of the 20 blocks needed to complete the design. Each square contained a colored dot. There were two boxes containing 100 1 × 1 × ½-inch wooden blocks. Each block was colored on the two square sides to match the colored dots.

Procedure

This experiment was conducted in a room at a day care center. *S* was seated at a small table with the puzzle board immediately in front of him. Four feet (from the center of the puzzle) to *S*'s immediate right was another small table upon which was placed one box of 100 blocks. Getting up, going to this box, and taking a block constituted a R_C. *E* was seated beside a table facing *S* at a distance of 6 feet (from the center of the puzzle) to the *S*'s immediate left. A second box of 100 blocks was on this table. In order to obtain blocks from this box, *S* had to leave his seat at the puzzle table and walk to *E*, who asked what color of block he wished. *S* answered and received a block from *E*. Acquiring a block in this way constituted an R_D.

E was a female who had previous experience as a child care worker. She invited *S* to play games and took *S* to the experimental room for the 15-minute trials.

Pretraining

Because pilot work had indicated a prominent R_C response set among 4-year-old children, a single pretraining trial was conducted in which potential *S*'s were trained to alternate R_D and R_C by means of prompting instructions and a concurrent FR 2 FR 2 schedule. Upon entering the experimental room for the pretraining trial, *E* explained the matching of the puzzle piece color to the colored dot on the puzzle board and said, "You can get the pieces for the puzzle from that box (points) by yourself. I will sit over there by that box (points) and you may get a piece from it by coming over and telling me what colored block you want. Then I will give it to you." She said, "O.K., now let's play the game. Remember, first you get a block from that box (points) and then you get a block from the other box (points). When you finish this puzzle, I will get a new one for you. You may start now." *S*'s were praised following every other dependent and every other competent response.

Pretesting

Pretraining was followed, on successive days, by two 15-minute pretesting trials without dependent or competent response contingent reinforcement.

Subjects

Five groups of 4-year-old children were then formed as follows: pretested *S*'s were rank ordered on the basis of their R_D/R_C ratio scores, bunched into groups of five on the basis of equality of their R_D/R_C ratio scores, and then one *S* from each bunch was randomly assigned to one of the five experimental groups. Each group contained five boys and three girls. Five subjects refused to participate in the training program and were replaced. Analysis of variance of the pretest R_D/R_C ratios of the experimental groups, following replacement, indicated no reliable differences.

The children's families were characterized by a very high divorce rate (65%) and lower middle class membership as measured by the Hollingshead Index of Social Position.

Training Trials

Pretesting preceded group assignment by 4–6 weeks. Following *S* selection and assignment, five 15-minute training, and four 15-minute extinction trials were administered on 9 consecutive work days. The experimental groups were as follows: Control, Double A, Double B, Dependent, and Competent. The reinforcement schedule for the two responses in

question was either Variable Ratio (VR) or Extinction (E). The concurrent schedules of reinforcement for the experimental groups were:

	R_C	R_D
Control	Ext.	Ext.
Double A	VR5	VR4
Double B	VR5	VR2
Dependent	Ext.	VR2
Competent	VR3	Ext.

The verbal communication from *E* to *S* during the experimental sessions consisted of: (1) puzzle completion contingent praise, and (2) R_D and/or R_C contingent praise.

RESULTS

The mean per trial R_D to total response rate ratios for the five experimental groups during training and extinction are reported for the sexes combined in Fig. 1.

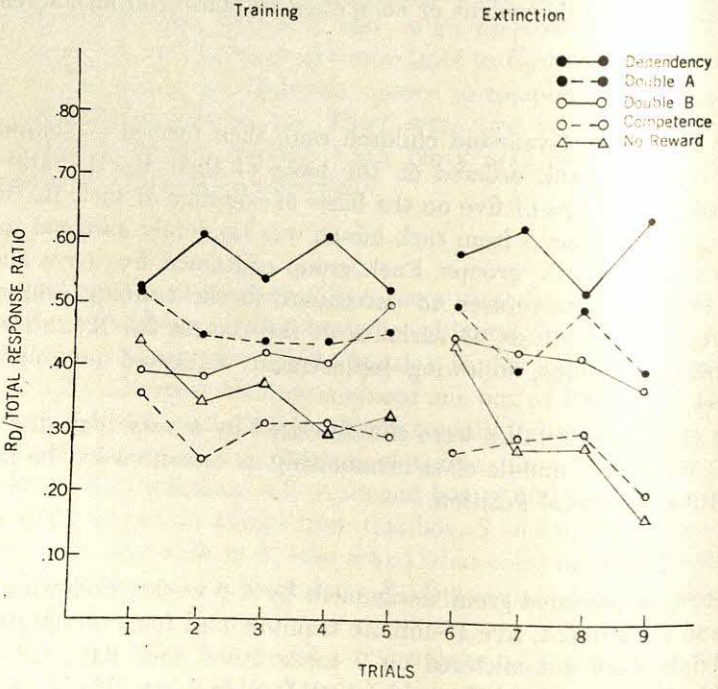


FIG. 1. Mean per trial R_D to total response ratios by experimental condition for sexes combined.

For the sexes combined, the high rate of R_D responding of the dependency group, the intermediate R_D rates of the two double rewards groups, and the lower R_D response rate of the control and competence reward group are apparent. Noteworthy is the increasing differentiation of the groups beginning during the later training trials and the marked separation of the samples during the last extinction trial.

Examination of the relative R_D response rates by sex (Figs. 2 and 3) suggests that the differentiation of the samples for the sexes combined was due in part to the high R_D rate of boys in the dependency reward group and in part to the low R_D rate of girls in the competence reinforcement sample.

However, examination of the R_C rate, R_D rate, and normalized ratio data using analysis of variance and t tests yielded only the following reliable group differences: (1) During training and during extinction the competence reward group had a higher mean R_C rate ($p < .05$) than did the dependency reward group or either double reinforcement sample; (2) during extinction girls emitted more R_D ($p < .05$) than did boys, independent of schedule of reinforcement. Table 1 presents these group means.

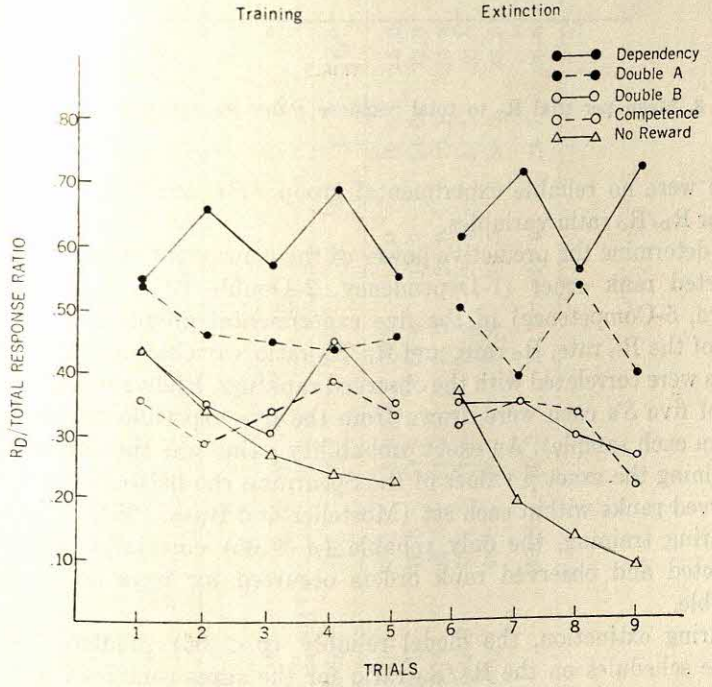


FIG. 2. Mean per trial R_D to total response ratios by experimental condition for boys.

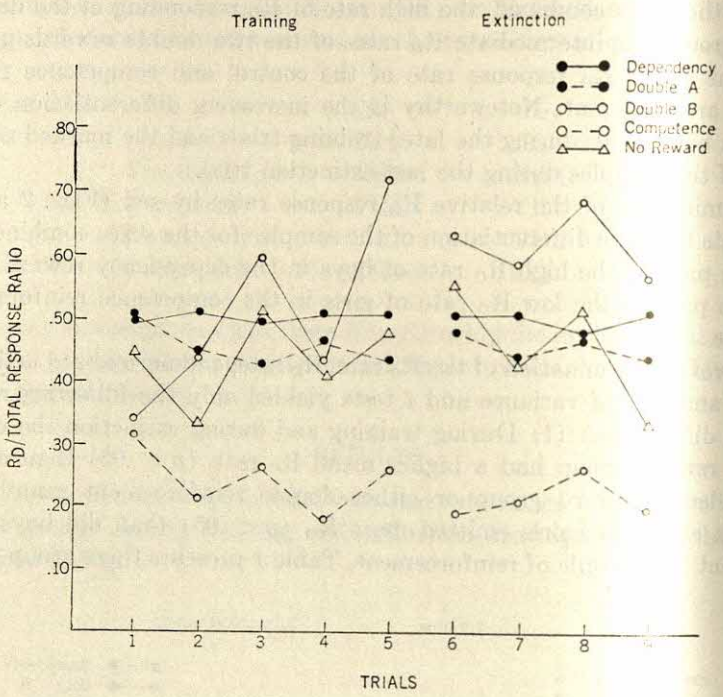


FIG. 3. Mean per trial R_D to total response ratios by experimental condition for girls.

There were no reliable experimental group differences on either the R_D rate or R_D/R_C ratio variables.

To determine the predictive power of the concurrent schedules model, a predicted rank order (1-Dependency, 2-Double B, 3-Double A, 4-No reward, 5-Competence) of the five experimental groups was derived on each of the R_D rate, R_C rate, and R_D/R_C ratio variables and the predicted orders were correlated with the observed rankings. Eight randomly selected sets of five *S*'s each were drawn from the five experimental groups (one *S* from each sample). An exact probability value was then determined by combining the exact p values of the Spearman rho between predicted and observed ranks within each set (Mosteller and Bush, 1954).

During training, the only reliable ($p < .05$) correlation between the predicted and observed rank orders occurred for boys on the R_C rate variable.

During extinction, the model reliably ($p < .05$) predicted the order of the schedules on the R_D/R_C ratio for the sexes combined and for the girls. The predicted and observed rank orders were reliably ($p < .05$) correlated on the R_C rate variable for the sexes combined and for boys during

TABLE 1

MEAN R_D/R_C RATIO T SCORES, R_D RATE AND R_C RATE BY SEX AND EXPERIMENTAL CONDITION

Experimental Condition	Training						Extinction					
	R_D/R_C T		R_D		R_C		R_D/R_C T		R_D		R_C	
	B	G	B	G	B	G	B	G	B	G	B	G
Dependent	58.1	53.9	19.8	25.0	19.8	25.0	59.2	53.7	15.8	20.2	11.3	20.5
Double-B	47.6	53.3	14.6	18.9	24.0	22.3	46.2	56.0	10.3	22.5	13.1	12.6
Double-A	53.3	51.9	15.2	21.0	19.1	26.3	51.2	51.6	12.2	19.5	15.3	24.2
No-Reward	45.1	51.5	13.2	20.4	24.5	26.0	44.9	51.2	5.6	16.5	18.5	26.3
Competent	45.0	40.7	16.4	12.5	36.0	33.6	45.1	43.3	13.1	11.9	31.9	34.0
Total	49.8	50.3	15.8	19.6	24.7	26.6	49.3	51.2	11.4	18.1	18.0	23.5

extinction. No reliable correlations were found on the R_D rate variables, for either sex, during training or extinction.

DISCUSSION

These findings provide varying degrees of support for the Gavalas and Briggs concurrent schedules of social reinforcement model of dependency. The correlational data suggests that the model is able to predict rank order effects moderately well. The relative absence of reliable parametric differences suggests that under the conditions of this study the model generated relatively poor behavior control and that there was considerable experimental group overlaps on the measures studied.

However, that the model predicted the rank order effects of the various schedules on the dependency-competence response ratio, and not on the rate of the dependent response variables, lends special emphasis to the behavioral relatively implicit in this definition of dependency. The facts that the model reliably predicted group orders on the R_D/R_C ratio and mean competent response rate, and *not* the mean rate of dependent responding, suggests that the concepts dependence and competence are not mutually exclusive. If a child's overall response rate is high, he may do considerable inefficient dependent socializing but still behave independently often enough to get things done.

That the relationship between the predicted and observed orders of the groups was stronger for girls than boys on the R_D/R_C ratio, and stronger for boys than girls on mean R_C, suggests a sex related asymmetry between dependence and competence at this age level.

Thus, preschool children who are socially rewarded for dependent behavior only can be expected to display more dependent behavior relative to competent behavior when the reinforcement is discontinued than children who are rewarded for both dependent and independent, task-relevant behavior. The latter children will tend to be more dependent than no-reward children, who will exceed children rewarded only for autonomous and efficient behavior on degree of dependent responding relative to competent responding.

In spite of this, the absence of reliable parametric differences between the experimental groups on the R_D/R_C ratio (the predictable measure of dependency) is important and warrants comment. It is possible that the influence of simple social reinforcement has been exaggerated. With the exception of the competent response rate of the competent-reward-only group, the rates of rewarded responses of all other experimental groups displayed a progressive and consistent decrease. As the experiment progressed it became apparent that the *S*'s were becoming bored with the task and were emitting considerable task irrelevant, novelty-seeking be-

havior. This is consistent with reports by Antonites, Frey, and Baron (1964), Long, Hammack, May, and Campbell (1958), and Long (1962 and 1963). Thus task satiation may have been a confounding variable in this study.

REFERENCES

ANTONITES, J. J., FREY, R. B., AND BARON, A. Group operant behavior: effects of tape recorded verbal reinforcers on the bar-pressing behavior of preschool children in a real life situation. *Journal of Genetic Psychology,* 1964, 1–5, 30–31.

BANDURA, A., AND WALTER, R. H. *Social learning and personality.* New York: Holt, 1963.

CATANIA, A. C. Concurrent operants. In W. K. Honig (Ed.), *Operant behavior: areas of research and application.* New York: Appleton, 1966.

FERSTER, C. B., AND SKINNER, B. F. *Schedules of reinforcement.* New York: Appleton, 1957.

GAVALAS, R. J., AND BRIGGS, P. F. Concurrent schedules of reinforcement: a new concept of dependency. *Merrill-Palmer Quarterly,* 1966, **12,** 97–121.

HARTUP, W. W. Dependency and independence. In H. W. Stevenson (Ed.), *Child psychology* The 62nd Yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press, 1963.

LONG, E. R. Additional techniques for producing multiple schedules control in children. *Journal of Experimental Analysis of Behavior,* 1962, **5,** 443–55.

LONG, E. R. Chained and tandum scheduling with children. *Journal of Experimental Analysis of Behavior,* 1963, **6,** 459–72.

MOSTELLER, F., AND BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.). *Handbook of Social Psychology.* Vol. I. Cambridge: Addison-Wesley, Readery, Massachusetts, 1954.

WALTERS, R. H., AND PARKS, R. D. Social motivation, dependency, and susceptibility to social influence. In L. Berkowitz (Ed.), *Advances in experimental social psychology,* Vol. I. New York: Academic Press, 1964.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 366–374 (1969)

Facilitation of Transposition and Reversal Learning in Children by Prior Perceptual Training[1]

THOMAS J. TIGHE AND LOUISE S. TIGHE
Dartmouth College

First-grade children who received either perceptual pretraining or the identical perceptual experience plus verbal labeling training were significantly and equally facilitated in performance on a discrimination reversal and a transposition test as compared to nonpretrained children. These results are discussed in terms of their implications for perceptual, attentional and verbal mediation theories of learning.

Kuenne (1946), and Alberts and Ehrenfreund (1951) found that the young child (below the age of about 5 years) is less likely than older *S*s to exhibit relational response in the simple transposition experiment, particularly with test stimuli remote from the training pair. Similarly, a number of studies have found that reversal is a very difficult task for the young child relative to other types of discrimination shifts but is a relatively easy problem for older children and adults (Kendler and Kendler, 1962; Tighe and Tighe, 1967). Efforts to explain such developmental differences in discrimination behavior have frequently drawn upon the concept of verbal mediation. Thus, an early hypothesis to account for transposition behavior is that as children mature they become more likely to utilize covert verbal mediating responses in the solution of such tasks, rather than simply responding to the physical properties of the stimulus objects *per se* (Kuenne, 1946). In the older child such verbalization (e.g., "the taller one is correct") is assumed to control and dominate choice behavior, thereby enabling the child to learn on a relational basis. Similarly, it has recently been proposed that relative ease of reversal shift by older children may be attributed to the development of dimensional control of discriminative response by means of such mediating representational responses (Kendler and Kendler, 1962). In support of

[1] This research was supported by Public Health Service Research Grant MH11088 from the National Institute of Mental Health. The authors are grateful for the generous cooperation of Gordon Tate, Superintendent of the Lebanon, New Hampshire, school system, and John A. Freitas, Superintendent of the Hartford, Vermont school system, and for the assistance of the principals and teachers of these school systems. We also wish to acknowledge the contribution of Mrs. Jane Higgens who assisted in collecting the data.

the verbal mediation hypothesis, it has been observed that giving children training in applying verbal labels descriptive of the relevant dimension speeds up reversal learning (Kendler and Kendler, 1961) and influences the direction of choice behavior in transposition tests (Marsh and Sherman, 1966). However, the notion that verbalization *per se* is the critical factor has been seriously questioned in recent reviews of reversal learning (Wolff, 1967) and transposition (Reese, 1968; Zeiler, 1967). In a comprehensive analysis of this issue Reese (1968) suggests that verbalization may facilitate reversal and transposition by mediating a perceptual response identifying or calling attention to the relevant dimension.

Differentiation theory (Gibson and Gibson, 1955; Tighe and Tighe, 1966), a recent theory of perceptual development, is consistent with these reviews in deemphasizing the role of verbalization in the development of discriminative behavior. This theory suggests that increased sensitivity to stimulus variables themselves may mediate the dimensional control which is assumed to be necessary for transposition and rapid reversal. According to this viewpoint, the abstraction of distinguishing features (dimensions) of stimulation is primarily a matter of learning through perceptual experience, and thus organisms may be expected to differ both individually and developmentally in their ability to detect stimulus dimensions and to independently utilize such invariant properties of stimulation to specify differences among objects. In support of this view is the recent finding that first-grade children were markedly facilitated in learning a discrimination reversal following pretraining of a perceptual nature which was designed to facilitate isolation of the task dimensions (Tighe, 1965).

The primary purpose of the present experiment is to test the hypothesis that a common perceptual learning process is involved in the development of transposition and reversal behavior. Secondarily, the experiment seeks an estimate of the significance of this process relative to verbal factors. The general procedure involves comparing the reversal and transposition performance of children given perceptual pretraining alone with that of children given identical perceptual exposure to the task stimuli but with added training in applying appropriate verbal labels to the stimuli.

METHOD

Subjects. The *S*s were 144 first-grade children ranging in age from 5 years 10 months–7 years 2 months with a mean age of 6 years and 7 months. This age was selected in view of evidence which suggests that such *S*s are at a transition stage in the development of discriminative processes (White, 1965). Each *S* was run individually in a small room adjacent to his classroom.

Pretraining procedure and apparatus. Subjects who received perceptual

pretraining only ($N = 48$) were required to make same–different judgments to successively presented stimulus objects which varied along the dimensions appearing in the subsequent discrimination task. The stimuli were selected from sixteen 1⅝-inch diameter wood cylinders which varied in height (Ht) and brightness (Br). Brightness was varied by covering the cylinders with black (B) or with white (W) enamel paint or with one of the following enamel mixtures—dark gray (one part W to four parts B) or light gray (one part B to 124 parts W). The Ht values were 4.5, 5.25, 6.75, and 7.5 inches. The 16 stimuli represented the possible combinations of these Ht and Br values. In order to aid the standardization of presentation conditions, each stimulus was placed in a 0.5-inch deep by 1.75-inch diameter hole cut in a 3.25-inch wide by 1-inch high plywood base or "stand" placed about 15 inches from *S* seated opposite *E* at a small table. The stimuli were placed into the plywood stand one at a time, and between presentations all stimuli were hidden from *S*'s view behind a nearby masonite screen. The total presentation was divided into eight subsets or series as described below. The first stimulus presented within a series constituted a standard (St) stimulus. The *S* was told, "Now let's pretend that this object is yours. Look at it carefully. I am going to take it away and show you some other objects one at a time. You tell me which of the other objects that you will see are exactly like yours." As each succeeding comparison (Co) stimulus was placed in the stand *E* said, "Is this object exactly like yours?" and the stimulus remained until *S* had rendered a judgment of "same" or "different." When an St was presented again to begin a new series, the first sentence of the instructions was changed to, "Here is your object again." When a new St was introduced the first sentence of the instructions was, "Now let's pretend that this object and only this one is yours."

The following stimuli were used as the standards: a B cylinder 5.25 inches high, a B cylinder 6.75 inches high, a W cylinder 5.25 inches high, and a W cylinder 6.75 inches high. Subjects underwent two series of Co stimulus presentations with each of the four Sts over a period of about 20 minutes. A series included the 16 possible combinations of the four Ht and four Br values, plus two interspersed reappearances of the standard cylinder.

The possible 24 orders of the four Sts were randomly distributed among all *S*s within the restriction that for half the *S*s the last St presented would be identical to one of the positive stimuli in the later discrimination and for the other half the last St would be identical to one of the negative stimuli. All judgments were based solely upon vision and *S*s received no information about correctness of judgment. Note that the general features of pretraining were such that *S*s were given a relatively

brief but concentrated experience with the fact that the objects could remain constant in an attribute (which could assume other values) while undergoing variation along another dimension, and they were given practice in discriminating the objects on the basis of one distinguishing feature independently of another.

Subjects who were given perceptual plus verbal pretraining ($N = 48$) received stimulus exposure and training *identical* to that given the perceptually pretrained *S*s but with added training in applying verbal labels to the stimuli. Whenever *S* said that the stimulus presented was not the same as his (i.e., as the standard), he was required by *E* to state whether the stimulus was taller, shorter, or the same as his and whether it was lighter, darker, or the same as his. On each occasion, then, *S* had to say two words, one to describe the height of the comparison stimulus relative to the standard and one to describe the brightness of the comparison stimulus relative to the standard. From the viewpoint of verbal mediation theory, it seems likely that such training should increase the availability of the appropriate mediating responses and thus should be an effective form of pretraining for these transfer tasks.

In order to make clear to *S* the meaning of the instructions he was to receive in pretraining, *S*s in the perceptual and perceptual plus verbal conditions were first given about 4 minutes practice in making corrected same–different judgments on cookie cutters varying in shape and hue.

An additional 48 *S*s were given pretraining on tasks unrelated to the subsequent discrimination tasks. This pretraining condition was designed to control for nonspecific transfer effects (e.g., total exposure to *E* and other aspects of the experimental situation). The tasks were modifications of the Picture Completion and Picture Arrangement subtests of the Wechsler Intelligence Scale for Children. The modifications consisted of (*a*) presenting all 20 pictures in the completion test to each *S* and (*b*) requiring each *S* to tell a story about each picture in the arrangement test. None of *S*'s responses in either task were corrected or reinforced.

Discrimination tasks. Immediately following pretraining all *S*s first learned a two-dimensional discrimination in which the heights of the choice objects were equal to the visible heights of the standard stimuli of pretraining, i.e., 4.75 and 6.25 inches. The stimuli were presented on a simple turntable device and on any one trial consisted of one of the following two pairs: (*a*) the tall white cylinder vs. the short black cylinder, or (*b*) the tall black cylinder vs. the short white cylinder. Tall and short were used equally often as the positive stimulus in all conditions of the experiment with the brightness dimension variable and irrelevant. The *S* was instructed to pick up one of the cylinders on each trial and was told that if his choice were correct he would find a marble under it.

The *S* was told that the "game" was to see how soon he could find a marble every time he chose. Training was continued in this manner until *S* made nine correct out of ten successive responses, at which point without change of instructions half of the *S*s in each pretraining treatment received a discrimination reversal while the other half received tests for transposition. Reversal training was continued to the same criterion as in the original discrimination. Transposition was tested by 20 presentations of a stimulus pair composed of either a 6.25- vs. a $8\frac{3}{16}$-inch cylinder ($N = \frac{1}{2}$ of the *S*s) or a 4.75- vs. a $3\frac{5}{8}$-inch cylinder ($N = \frac{1}{2}$ of the *S*s). All choices were reinforced. During both the transposition test and reversal learning the irrelevant dimension of brightness continued to vary. Each *S* was assigned to groups and to positive stimuli on a predetermined random basis. At the end of the experiment *S* was allowed to choose a prize from an assortment which included charms, bubble gum, whistles, raisins, jack sets, Tootsie rolls, pencils, pencil sharpeners, M and Ms, and flutes.

RESULTS

Pretraining. Differences between the treatment groups to be compared on the transfer tasks were evaluated by *t* tests. Among *S*s who underwent reversal learning, the mean number of errors in pretraining were 19.1 for perceptually pretrained *S*s and 15.1 for those given perceptual plus verbal pretraining, $t = 1.34$, $df = 46$, $p > .10$. For *S*s given transposition tests the corresponding means were 18.7 and 15.0, $t = 1.40$, $df = 46$, $p > .10$. For *S*s who underwent reversal learning, the mean number of errors during the final series of pretraining judgments were 2.12 for perceptually pretrained *S*s and 1.58 for *S*s given perceptual plus verbal pretraining, $t = 1.23$, $df = 46$, $p > .20$. For *S*s given transposition tests the corresponding values were 1.63 and 2.04, $t = 1.07$, $df = 46$, $p > .20$.

Initial discrimination. The mean number of trials to criterion for *S*s who underwent perceptual, perceptual plus verbal, and control pretraining prior to reversal learning were 27.1, 21.5, and 29.3, respectively. For *S*s given transposition tests the corresponding means were 44.8, 35.8, and 25.0. Analyses of variance on this measure yielded $F = .47$, $df = 2,69$, and $F = 2.21$, $df = 2,69$, $p > .10$, for the three reversal groups and the three transposition groups, respectively. Analyses of errors produced similar conclusions.

Thus, the major treatment groups did not differ significantly in performance in pretraining or in the initial discrimination.

Reversal learning. The mean number of trials to criterion in the discrimination reversal was 9.46 for the group given perceptual pretraining, 15.75 for the group given perceptual plus verbal pretraining, and 40.21

for the group pretrained on the control tasks. A $\sqrt{X + .5}$ transformation was applied to the trials to criterion measure and a 3×2 analysis of variance was carried out (type of pretraining by tall vs. short positive). Type of pretraining was a significant source of variance ($F = 8.48$, $df = 2,66$, $p < .001$), while neither stimulus assignment nor the interaction approached significance (both F values < 1). The control *Ss* reversed significantly more slowly than both the *Ss* given perceptual pretraining ($t = 3.88$, $df = 46$, $p < .001$) and the *Ss* given perceptual plus verbal pretraining ($t = 2.75$, $df = 46$, $p < .01$). The latter two groups did not differ in speed of reversal learning ($t = 1.10$, $df = 46$, $p > .20$). The same conclusions result from analyses of errors.

Transposition. In analyzing the data of the transposition test, an *S* was considered to have exhibited transposition if he made 16 or more choices of a relational nature during the 20 transposition trials—a choice pattern expected to occur by chance less than one time in 100. The number of *Ss* making 16 or more relational choices in the groups given perceptual pretraining, perceptual plus verbal pretraining, and the control tasks were 21, 18, and 10, respectively, out of a possible 24 *Ss* in each group. The control *Ss* differed significantly on this measure from both perceptually pretrained *Ss* ($\chi^2 = 9.11$, $p < .005$) and *Ss* given perceptual plus verbal pretraining ($\chi^2 = 4.20$, $p < .05$). The latter two groups did not differ ($\chi^2 = 0.55$, $p > .30$). The same pattern of conclusions resulted from comparisons of the frequency of transposition (relational) responses within each treatment. The mean number of transposition responses in the perceptual, perceptual plus verbal, and control conditions were 18.5, 16.8, and 13.7, respectively. By Wilcoxon's test for unpaired replicates the controls differed on this measure from both perceptually pretrained *Ss* ($p < .01$) and *Ss* given perceptual plus verbal pretraining ($p < .05$), while comparison of the latter two groups did not approach significance.

It might be noted that the control *Ss* exhibited less transposition than is usually observed in *Ss* of this age on "near" tests. This outcome is probably due to the presence of the variable irrelevant dimension in the present transposition experiment.

Finally, a square root transformation was applied to the number of transposition responses by each *S* and a $3 \times 2 \times 2$ analysis of variance was carried out (type of pretraining by transposition upward vs. downward by former positive vs. former negative stimulus in the transposition test pair). Although the distributions of transposition responses were so markedly skewed as to render questionable the appropriateness of such an analysis, it was nevertheless deemed advisable to employ a statistical test which would check the possible interactions among these factors. Only type of pretraining proved to be a significant source of variance

($F = 5.17$, $df = 2,60$, $p < .01$). The F values associated with the other main effects and with the interactions did not approach significance.

DISCUSSION

These results establish that discrimination reversal learning and transposition can be markedly facilitated by a common training procedure which is essentially perceptual in nature. As such, the data are in agreement with the conceptions of the differentiation theory of discrimination learning (Tighe and Tighe, 1966, 1968a), which relate developmental changes in discrimination behavior to changes in *S*'s sensitivity to stimulation. From this viewpoint, the facilitation observed in both transposition and reversal is seen as resulting from an increase, through perceptual pretraining, in *S*'s ability to detect the relevant dimension-reward relation. In contrast, non-pretrained *S*s of this age are assumed to be more likely to accomplish the initial discrimination and subsequent transfer tasks on the basis of the specific stimulus object-reward relations in each problem. Only when *S* has fully differentiated out the common feature-reward relation in these tasks is there a basis for transfer of relational response in the transposition test or for positive transfer to reversal in the form of attention to the relevant feature-reward relation. Alternatively, attentional theories of discrimination learning (e.g., Sutherland, 1959; Zeaman and House, 1963) might deduce the pattern of results from the assumption that pretraining serves to focus *S*'s attention on the task dimensions vis-à-vis other features in the general experimental situation. But from either theoretical viewpoint it is surprising that perceptual pretraining did not facilitate performance in the initial discrimination as might be expected under the assumption that pretraining increased dimensional control over discriminative response. Note, however, that the initial discrimination, unlike reversal or transposition, requires considerable learning in addition to the detection of the relevant dimension. That is, it is in the initial discrimination that *S*s must learn the "rules of the game," and it is possible that pretraining did not constitute an appreciable advantage in the face of such nonspecific learning (e.g., the elimination of position or sequence hypotheses). That such requirements are a significant factor in the learning of these tasks is attested to by Jeffrey's study (1965) which found a 14-fold reduction in the mean number of trials to criterion in the learning of a similar task by four year olds who had been pretrained in a series of object-quality discriminations.

The finding that perceptual plus verbal pretraining was not more effective than perceptual pretraining alone (and, in fact, tended to be somewhat less effective) is also surprising in view of the widely held assumption that differences in the strength or availability of appropriate

verbal mediators play a major role in the developmental phenomena of discrimination. Note that this result cannot be attributed to a differential effect of the pretraining treatments on performance during pretraining or the initial discrimination. And it is unlikely that this outcome is due to a "ceiling effect" of perceptual pretraining alone, at least in regard to reversal learning, since considerably more rapid reversal has been repeatedly observed with special training in similarly structured tasks with *Ss* of the same age (e.g., Tighe and Tighe, 1965; Tighe and Tighe, 1968b). Rather, it appears that overt verbalization training simply did not constitute an added advantage to performance in these tasks beyond that provided by the conditions of perceptual pretraining alone. This observation suggests that some of the beneficial effects upon discrimination which have been attributed to verbal pretraining procedures may, in fact, be due primarily to the perceptual experience which is an inevitable concomitant of such pretraining. This consideration underscores the caution that studies investigating transfer to discrimination from verbal pretraining should include a control condition in which *Ss* make prior differential perceptual responses to the stimulus elements.

Although we have hypothesized that the critical changes which took place in pretraining were of a perceptual nature, it is of course possible that *Ss* in the perceptual pretraining group were covertly using labels or other mediating representational responses and, therefore, it was this practice, rather than perceptual learning, which accounts for the observed facilitation. However, it seems unlikely that such mediators played a critical role when one considers the nature of the conditions which have been found to be essential for effective perceptual pretraining (Tighe and Tighe, 1968b, 1968c). The latter experiments show that perceptual pretraining is effective in facilitating discrimination only under a specific set of conditions (which were employed in the present experiment) while a variety of other conditions, which appear to allow equal opportunity for covert mediation, have no effect. For example, perceptual pretraining with either three or four stimulus values appearing on each dimension facilitates later reversal learning, while perceptual pretraining with two stimulus values per dimension has no effect on later learning. This observation is consistent with the assumptions of differentiation theory (Tighe and Tighe, 1968b), but is unexpected from the viewpoint of verbal mediation theory.

REFERENCES

ALBERTS, E., AND EHRENFREUND, D. Transposition in children as a function of age. *Journal of Experimental Psychology*, 1951, **41**, 30–38.

GIBSON, J. J., AND GIBSON, E. J. Perceptual learning: Differentiation or enrichment? *Psychological Review*, 1955, **62**, 32–41.

JEFFREY, W. E. Variables affecting reversal-shifts in young children. *American Journal of Psychology,* 1965, **78,** 589–595.

KENDLER, H. H., AND KENDLER, T. S. Effect of verbalization on reversal shifts in children. *Science,* 1961, **134,** 1619–1620.

KENDLER, H. H., AND KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychological Review,* 1962, **69,** 1–16.

KUENNE, M. R. Experimental investigation of the relation of language to transposition behavior in young children. *Journal of Experimental Psychology,* 1946, **36,** 471–490.

MARSH, G., AND SHERMAN, M. Verbal mediation of transposition as a function of age level. *Journal of Experimental Child Psychology,* 1966, **4,** 90–98.

REESE, H. W. *The perception of stimulus relations: Discrimination learning and transposition.* New York: Academic Press, 1968.

SUTHERLAND, N. S. Stimulus analyzing mechanisms. In *Proceedings of a symposium on the mechanization of thought processes.* Vol. 2. London: Her Majesty's Stationery Office, 1959. Pp. 575–609.

TIGHE, L. S. Effect of perceptual pretraining on reversal and nonreversal shifts. *Journal of Experimental Psychology,* 1965, **70,** 379–385.

TIGHE, L. S., AND TIGHE, T. J. Overtraining and discrimination shift behavior in children. *Psychonomic Science,* 1965, **2,** 365–366.

TIGHE, L. S., AND TIGHE, T. J. Discrimination learning: Two views in historical perspective. *Psychological Bulletin,* 1966, **66,** 353–370.

TIGHE, L. S., AND TIGHE, T. J. Transfer from perceptual pretraining as a function of number of stimulus values per dimension. *Psychonomic Science.* 1968c, **12,** 135–136.

TIGHE, T. J., AND TIGHE, L. S. Discrimination shift performance of children as a function of age and shift procedure. *Journal of Experimental Psychology,* 1967, **74,** 466–470.

TIGHE, T. J., AND TIGHE, L. S. Perceptual learning in the discrimination processes of children: An analysis of five variables in perceptual pretraining. *Journal of Experimental Psychology,* 1968b, **77,** 125–134.

TIGHE, T. J., AND TIGHE, L. S. Differentiation theory and concept-shift behavior. *Psychological Bulletin,* 1968a, **70,** 756–761.

WHITE, S. H. Evidence for a hierarchical arrangement of learning processes. In L. P. Lipsett & C. Spiker (Eds.), *Advances in child development and behavior.* Vol. 2. New York: Academic Press, 1965. Pp. 187–220.

WOLFF, J. L. Concept-shift and discrimination-reversal learning in humans. *Psychological Bulletin,* 1967, **68,** 369–408.

ZEAMAN, D., AND HOUSE, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency.* New York: McGraw-Hill, 1963. Pp. 159–223.

ZEILER, M. D. Stimulus definition and choice. In L. P. Lipsett and C. Spiker (Eds.), *Advances in child development and behavior.* Vol. 3. New York: Academic Press, 1967. Pp. 125–156.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 375–388 (1969)

The Child's Memory for Unitized Scenes[1]

LEONARD M. HOROWITZ, ANITA K. LAMPEL, AND RUBY N. TAKANISHI[2]
Stanford University

This article shows that the nursery school child remembers a set of pictures or objects better if the items are unitized. For example, a shoe, key, and box are better remembered if the key is in the shoe and the shoe is in the box. Even arbitrary units are better remembered than a series of items. In Exp. I, the child was first shown a display of pictures or objects. He named each item and the items were then hidden; then one was removed and the remaining items were exposed. The *S* was then asked what was missing. Different groups of *S*s were tested in different conditions. The unitized condition showed a clear superiority. Experiment II replicated this result with a simpler procedure and with each *S* as his own control. Experiments III and IV then examined the role of verbal descriptions and verbalization in producing the effect.

When does a collection of objects form a "unit"? If a child is simultaneously shown "a shoe, a key, and a box," how can the objects be unitized into a single display? The present article shows that *E* can make a unit of the objects by placing the key in the shoe and the shoe in the box. When the nursery school child recalls this scene, he seems to remember it as a whole. This article shows, though, that the objects are not as well remembered if they are simply lined in a row.

Recent research has examined this issue in some detail. Rohwer and his associates (e.g., Suzuki and Rohwer, 1968) and Reese (1965) have shown how much better two objects are remembered if the objects interact in some way. Apparently, a relationship between two objects integrates them into a unit. However, we still do not understand what conditions "unitize" a scene. Is *any* arbitrary relationship able to unitize the scene? If so, why can't two objects be unitized by the relationship "A is next to B"? If we understood the properties of a unit, we might be able to predict how and when a unit is formed.

Units and redintegrative memory. In general, tasks of memory are conveniently analyzed into stimuli eliciting responses. In some tasks,

[1]This research was supported in part by Grant GB-8365 of the National Science Foundation to the senior author. The authors are indebted to the National Science Foundation for this support.

[2]Now at the University of Chicago.

though, the stimulus is itself a part of the response. When the *S* sees this stimulus, he thinks of the whole scene of which it was a part. Seeing an arrow, he does not think "bow"; instead, he remembers: "the bow and arrow that I saw in the hands of the skillful archer yesterday." This kind of memory has been called "redintegrative" memory (Hollingworth, 1928). It seems to be related to the memory for units.

A simple version of redintegrative memory has recently been studied in adult *S*s by Horowitz and Prytulak (1969). The *S* was asked to recall a list of items—words, phrases, sentences, or 2-digit numbers—by the method of free recall. The memory was called "redintegrative" if it met this criterion: Whenever *S* recalled any part of an item, the probability was very high that he would recall the whole item. For example, recalling an adjective-noun phrase is a redintegration: An *S* rarely recalls a lone adjective or a lone noun; if he recalls either part, the chances are high that he will recall the whole phrase. Thus, the adjective-noun phrase seems to be a unit.

Tasks that meet this criterion have an interesting property. This property appears when *free* recall is compared with *cued* recall: Suppose an *S* is remembering phrases like *heavy cake, dry hair,* etc. In free recall he usually remembers the whole phrase, of course, but there are times when he only remembers a part; for adjective-noun phrases, the remembered part is more apt to be the noun. On the other hand, suppose *S* is given a hint about a phrase. Which part of the unit is the better cue, the adjective or the noun? Horowitz and Prytulak showed that the noun is also the better cue. Thus, the part which is more probable in free recall is more effective in eliciting the whole.

In general, when we remember units, one part of the whole unit is more salient. When other parts are forgotten, this part might still be remembered. This important part is also said to have more "redintegrative power" than other parts; it is a better cue for eliciting the whole.

Thus, a unit might be characterized this way: A unit has an organization; the different parts vary in their importance to the whole. One part is more important or more salient. That part is more apt to be remembered, and once it is remembered, it elicits the whole unit more effectively.

Now suppose a unitized scene is composed of four ingredient objects, A, B, C, and D, and let us assume that *S* perceived each of these ingredients. Furthermore, let us say that B is the single most important ingredient in the whole; no single cue would be more effective than B. However, several parts could be combined to yield a larger fraction of the whole. "A + B" contains more of the whole than A alone, so "A + B" ought to have more redintegrative power than either A or B alone. And likewise,

"A + B + D" should elicit the whole more strongly than "A + B." Therefore, the cue "A + B + D" should elicit the whole unit extremely well, and the *S* should easily recall the missing ingredient C.

Now consider the following experimental procedure. Suppose *S* sees a unitized scene and the *E* then removes one ingredient part—say, Part C. Then *S* is asked what is missing. If the scene were unitized, the remaining "A + B + D" should readily elicit the whole, and *S* should easily recall the missing part. But if the scene were not unitized, the missing object would not be as readily recalled. In fact, for non-unitized scenes, A, B, and D might each elicit irrelevant associates which could interfere with *S*'s memory for the missing C. Thus, a test of the missing part should readily distinguish between units and non-units.

The present studies demonstrate that the missing part of a unitized display is almost perfectly remembered. Memory is very much poorer if the same objects are not unitized. These studies also examine the conditions that produce the difference and therefore help determine what makes a "unit."

EXPERIMENT I

Method

Test materials. Each child was tested on 14 displays. There were two practice displays, six picture displays, and six object displays (in that order). Each practice display contained two objects, each picture display contained two pictures, and each object display contained three objects. The objects were all small toys like those found in the Stanford-Binet test (cat, train, doll); they varied in size and color. The pictures were black-and-white line drawings of simple objects.

The practice displays were: (1) spoon and cat; (2) boat and button. The picture displays were: (1) pail and duck; (2) book and scissors; (3) foot and cup; (4) squirrel and umbrella; (5) apple and tree; (6) rabbit and wagon. The object displays were: (1) doll, chair, and ball; (2) horse, flag, and train; (3) whistle, shoe, and box; (4) monkey, bottle, and television set; (5) block, shoestring, and key; (6) car, dog, and ring.

Each display was presented either as a series or as a unit. In the unitized condition (Cond. U) the ingredients of each display were arbitrarily integrated into a scene. Sometimes the scene was realistic and other times it was not. For example, in the first object display, the doll sat on the chair with the ball in her lap; in the second object display, the horse stood on the train with the flag leaning out of the window. Realistic and

unrealistic scenes did not differ systematically, so this variable will not be discussed further.[3]

The other two conditions were called serialized (S) conditions; one was called "the simultaneous Cond. S" and the other "the temporal Cond. S." In these two conditions the objects or pictures of each display were shown in a series. In the simultaneous Cond. S the ingredients appeared in a row before the subject. Adjacent items were about 1/2 inch apart. In the temporal Cond. S, the ingredients appeared successively one after another on the very same spot.

Procedure. Each *S* was tested individually. The *S* entered the "surprise room" of the nursery school and was seated opposite *E* at a small table. On the table before him there were two brightly colored, cylindrical clowns made of construction paper. The clown was 12 inches high and its mouth was large enough to accommodate an object or picture. Anything inserted into the clown's mouth fell out of *S*'s sight. There was a slit in the back of the clown, so *E* could remove the object without *S* seeing it.

E said: "We are going to play the 'What's Missing Game' with these clowns." Then *E* presented the first practice display: She placed the toy spoon on the table before *S* and asked "What's this? What is it called?" After *S* named the object, *E* dropped it into one clown's mouth. Then she placed the toy cat before *S*, asked him to name it, and then dropped it into the clown's mouth. Finally, *E* told *S* to close his eyes, and she removed the cat from the slit in the back.

E then told *S* to open his eyes, and she recited a short poem, urging *S* to join in and clapping her hands in rhythm with the poem: "This is the way we play the game. What is missing? What's its name?" Then the two clowns were lifted simultaneously to expose the remaining object (the spoon). Finally, *E* asked *S* what was missing. The total delay from exposure to test lasted about 20 sec. The procedure was then repeated with the second practice display.

Then the child was tested with the six picture displays and after that, with the six object displays. The order of items was the same for all *S*s. In Cond. U and in the simultaneous Cond. S, all ingredients of the display appeared together. *E* pointed to the ingredients one by one and asked *S* to name each one. In the temporal Cond. S, *E* presented each ingredient singly: She presented one ingredient, asked *S* to name it, dropped it into the clown's mouth, and then presented the next ingredient.

Each ingredient of a display was inserted into a different clown's

[3] As described below, every *S* in Cond. U showed nearly perfect performance. Thus, the ceiling effect in this condition may have obscured a slight difference between realistic and unrealistic scenes. One point, however, is clear: A scene can be unitized without being realistic

mouth. This rule held for all experimental conditions. When a display contained three objects, a third clown was added to the table.

The missing ingredient of each display was systematically varied from one subject to the next. In each test *S* was allowed 20 sec to respond. His score told how often he responded correctly.

Subjects. The *S*s were 72 pupils at the Bing Nursery School of Stanford University. Each group contained an equal number of males and females. Two age levels were studied; the 36 younger *S*s ranged from 3 years 10 months to 4 years 5 months. The 36 older *S*s ranged from 4 years 8 months to 5 years 3 months.

Results and Discussion

Table 1 reports the mean number of correct responses to the picture displays. The label "_X" tells that the first picture of a series was missing in the test, and "X_" tells that the second picture was missing. The maximum score for each position of recall was 3.0.

TABLE 1

MEAN NUMBER OF CORRECT RECALLS, PICTURE DISPLAYS (MAX. SCORE = 3)

	Unitized		Serialized (simultaneous)		Serialized (temporal)	
	_X	X_	_X	X_	_X	X_
Age 3.10–4.05	3.00	3.00	2.00	1.92	1.58	1.25
Age 4.08–5.03	3.00	2.92	2.25	2.25	1.83	2.08

The data were not normally distributed, nor did they display homogeneity of variance. For heuristic reasons, however, the authors performed an analysis of variance, adopting the .01 level of significance. (Several authors have recommended this kind of approach—e.g., Lindquist, 1953.) Then the conclusions of that analysis were confirmed by a chi-square test.

The analysis of variance was performed on *S*'s total score. The total score could range from 0 to 6. This analysis showed that the difference between conditions was highly significant, $F\ (2,66) = 49.67$, $p < .001$. No other main effect or interaction was significant. In this and subsequent experiments, female *S*s performed better than male *S*s, but the difference did not reach statistical significance. Nor did sex interact with other variables. Therefore, the effect of sex will not be discussed further.

To examine the two positions of recall, a separate analysis of variance was performed within each experimental condition. There was no point, of course, in analyzing the data of Cond. U. In the other two conditions, no source of variance reached statistical significance; all F values < 1.81.

Data for inidividual *S*s showed equal recall for "_X" and "X_." In Cond. U, 23 *S*s showed identical scores, 1 *S* recalled more items to the cue "_X," and no *S* recalled more items to the cue "X_." For the simultaneous Cond. S, the corresponding figures were 8, 8, and 8. For the temporal Cond. S, the corresponding figures were 3, 10, and 11. Thus, pictures were recalled equally well to either kind of cue.

Then a chi-square test was performed to confirm all of these conclusions. The partitioning of a chi-square has been discussed by Winer (1962), Sutcliffe (1957), and Wilson (1956). To perform this test, a contingency table was prepared. Two columns of the contingency table reported whether *S*'s score was above the median or below the median. The rows reported the different combinations of the three variables—the three experimental conditions, the two age levels, and the two positions of recall. Observed frequencies were entered in each cell; each expected frequency exceeded five.

The overall value of $X^2(11df)$ was 75.66, $p < .001$. Following the procedure of Sutcliffe (1957) and Wilson (1956), this chi-square was then partitioned into its components. Only one component reached statistical significance; $X^2_{\text{Conditions}}(2\ df) = 70.74$, $p > .001$. All other components were not significant, $p > .05$. This analysis, then, agreed perfectly with the analysis of variance.

To summarize, the picture displays of Cond. U were best remembered and those of the temporal Cond. S were most poorly remembered. Furthermore, recall was equally good in both directions. Many cases of symmetrical recall have been found in adults (e.g., Ekstrand, 1966; Horowitz, Norman, and Day, 1966), so it is interesting to find symmetrical recalls in these data on children.

Table 2 reports the mean number of correct responses to the object displays. Notations like "_X X" tell the position of the missing object. The maximum score at each position was 2.00.

Again, as a first step, an analysis of variance was performed using each *S*'s total score. The total score could range from 0 to 6. The results were

TABLE 2

MEAN NUMBER OF CORRECT RECALLS, OBJECT DISPLAYS (MAX. SCORE = 2)

	Unitized			Serialized (simultaneous)			Serialized (temporal)		
	_XX	X_X	XX_	_XX	X_X	XX_	_XX	X_X	XX_
Age 3.10–4.05	1.83	2.00	2.00	1.08	0.83	1.50	0.58	0.67	1.08
Age 4.08–5.03	2.00	2.00	2.00	1.67	1.17	1.50	1.00	0.50	0.92

very similar to those of Table 1. The difference among conditions was significant, $F(2,66) = 91.27$, $p < .001$; but no other source of variance was significant. The unitized condition produced nearly perfect recall at both age levels. The simultaneous Cond. S was poorer, and the temporal Cond. S, worst of all. This result was also confirmed by a chi-square test like that reported above. The overall value of X^2 (17 *df*) = 122.0; $X^2_{\text{Conditions}}(2\ df) = 103.44$, $p < .001$.

To examine the effect of the item's position, a separate analysis was performed on the data of each condition. Again, there was no point in analyzing Cond. U further. However, the three positions of recall did differ significantly in the simultaneous Cond. S ($F = 4.37$, $p < .05$) and in the temporal Cond. S ($F = 2.67$, $p < .10$).

Since the three positions differed significantly in both versions of Cond. S, the serial position effects are worth examining further. In both cases, the overall recall was best for the item at Position 3 and poorest for the item at Position 2. This recency effect seems to characterize children's memory (e.g., Hagen and Kingsley, 1968).

The recency effect was particularly evident in the data of the younger *Ss*. The older *Ss* seemed to show a slightly superior recall at Position 1. This trend towards a *primacy* effect in the older child implies an interaction between age and position. According to the chi-square analysis, this interaction reached marginal significance: $X^2_{\text{Age} \times \text{Position}}(2\ df) = 5.15$, $p < .10$.

Memory for words and intact language typically exhibits a primacy effect with adults: The beginning of the sequence is better remembered, and the middle is most poorly remembered. It is particularly interesting in the present data that the older *Ss* displayed a primacy effect. This result is related to a result of Marchbanks and Levin (1965). These authors showed that kindergarten children respond more strongly to the initial letter of a nonsense word than to any other letter. Thus, a trend towards a primacy effect may develop as a child approaches the age for reading. This point needs to be explored more fully.

EXPERIMENT II

Experiment 1 suggested an enormous difference between the unitized and serialized conditions. This result was obtained through a "between-subjects" experimental design: Every *S* served in one condition only.

Experiment 2 was designed to extend this result to a "within-subjects" or "mixed-list" experimental design. From studies with adults, the two procedures do not always yield equivalent results (Twedt and Underwood, 1959).

Method

Experiment 2 was conducted by a different *E*. Every *S* was tested in two experimental conditions—Cond. U and the simultaneous Cond. S. To shorten the task a little, two picture displays were eliminated (book and scissors, apple and tree), leaving two practice displays, four picture displays, and six object displays. Two picture displays and three object displays were presented under Cond. U, while the remaining displays were presented under Cond. S. Each particular display was unitized for half the *S*s and serialized for the other *S*s.

To make the procedure more efficient, there were several other minor changes. First, only one clown was used; the two or three ingredients of a display were dropped into the mouth of that single clown. Second, after each ingredient was concealed, *E* did not recite the poem and clap her hands. Instead, she told *S* to close his eyes, then she removed one ingredient and told *S* to open his eyes; finally she lifted the clown asking "What's missing?" The procedure was thus streamlined, and the retention interval also became a little shorter. As shown below, however, these changes made no difference.

The *S*s were 12 pupils (six males and six females) from the Bing Nursery School of Stanford University. They ranged in age from 3 years, 6 months, to 3 years, 11 months.

Results and Discussion

The mean number of correct responses in each condition is reported in Table 3. Both sets of material showed better performance in Cond. U than in Cond. S.

TABLE 3
MEAN NUMBER OF CORRECT RESPONSES

	Cond. U	Cond. S
Pictures (Max. score = 2)	2.00	1.42
Objects (Max. score = 2)	2.83	2.08

A difference-score (Cond. U.–Cond. S) was computed from the data of each *S*. An analysis of variance was performed on the picture displays and on the object displays. No sources of variance were significant for either type of material; all F values < 2.26, $p > .05$. Then the error term of the analysis of variance was used to test whether the mean difference-

score differed significantly from 0. In both cases, the difference was significant; t (8) > 3.44, $p < .02$.

The data of individual *S*s clearly showed the advantage of unitized over serialized displays. For the data on pictures, six *S*s performed better on U items, none performed better on S items, and six performed equally well on both types. For the data on objects, seven *S*s performed better on U items, one performed better on S items, and four performed equally well on both types. Eleven of the 12 *S*s showed an advantage for U items on at least one type of material.

The results of Exp. I can therefore be extended. The experimental effect continued to hold (a) when the procedure was streamlined in several ways, and (b) when each subject served in both conditions. Since an *S* can serve as his own control, there cannot be much transfer from one display to the next: After a child sees several unitized displays, he does not seem to unitize a serial display spontaneously. If he did, the difference between conditions would have diminished.

EXPERIMENT III

What produces the difference between serialized and unitized displays? Perhaps the advantage of a unitized display can be linked to the verbal process. A child may translate each scene into words and largely remember his own words. If so, a child's verbal description of a unitized scene should differ from his verbal description of a serialized display.

Unitized displays show objects interacting. An adult's description of a unitized display would therefore involve verbs and prepositions. Perhaps these verbs and prepositions enhance *S*'s memory for unitized displays. If so, *S*'s descriptions of unitized displays should contain more words, and these extra words should be mainly verbs and prepositions.

Several experiments (e.g., Suzuki and Rohwer, 1968) have already shown that verbs and prepositions enhance an older child's memory for noun pairs. An *S* can remember "car-wagon" much more readily after hearing "The car *pulled* the wagon" than after hearing "The car *and* the wagon."

We therefore hypothesized that even the youngest *S*s would use more words in describing the unitized displays. Their descriptions were expected to be longer, mainly because of the verbs and prepositions. This hypothesis was tested in Exp. III.

Method

Each *S* was shown the display of Exp. II and asked to describe what he saw. The *E* said: "What's this? Tell me about what you see." The *E* recorded *S*'s description verbatim. Half of the items were unitized, and

half were serialized. The *E* scored the total number of words that *S* produced, and also the number of verbs and prepositions. Twelve *Ss* were tested. Half were males and half were females; half ranged from 3 1/2 years to 4 years, and half ranged from 4 1/2 years to 5 years.

Results and Discussion

Table 4 reports the mean number of words *S* produced in describing the displays. The words "the," "a," and "and" were not counted in this score.

A single analysis of variance was performed on these data (Lindquist, 1953, "Type VI"). This type of analysis yields three "within-subjects" error terms. The Bartlett test of homogeneity of variance showed that the three error terms did not differ significantly ($X^2 = 0.260$, $p > .10$), so the three terms were pooled to yield a single, more stable error term with 30 *df*. The analysis of variance showed several significant sources of variance. For one thing, the objects elicited many more words than the pictures; $F\ (1,30) = 74.98$, $p < .001$. There was also a significant age effect; $F\ (1,10) = 12.82$, $p < .01$. More important, though, age interacted significantly with other variables: First, age interacted with the experimental condition (S vs. U), $F\ (1,30) = 4.32$, $p < .05$. And second, age interacted with the nature of the material (pictures vs. objects), $F\ (1,30) = 5.47$, $p < .05$.

The interactions with age are due to a smaller difference for younger *Ss*. The interaction between age and experimental condition was especially interesting: Unitized scenes elicited more words than serialized scenes for the older *Ss*; but the younger *Ss* showed no difference at all. For the younger child, a unitized display does facilitate memory, but it does not elicit any more words of description.

Verbs and prepositions are considered to be important in helping a child unitize a cognition (e.g., Suzuki and Rohwer, 1968). Therefore, the numbers of verbs and prepositions emitted by each subject were counted.

TABLE 4
TOTAL NUMBER OF WORDS EMITTED (Excluding "the," "a," and "and") PER ITEM PER CHILD

	Cond. U	Cond. S
Age 3 1/2–4		
Pictures	4.0	4.3
Objects	8.0	8.7
Age 4 1/2–5		
Pictures	6.5	4.5
Objects	14.1	11.4

TABLE 5

TOTAL NUMBER OF VERBS AND PREPOSITIONS PRODUCED PER CHILD

	Cond. U	Cond. S
Pictures		
Age 3 1/2–4	0.2	0.0
Age 4 1/2–5	1.7	0.0
Objects		
Age 3 1/2–4	0.2	0.0
Age 4 1/2–5	3.5	0.0

Table 5 reports the mean scores. The data were not analyzed statistically because of the large number of 0's, but the message is very clear: Older *Ss* spontaneously produced verbs and prepositions in describing the unitized displays, while younger *Ss* did not. Although verbs and prepositions do exist in the vocabularies of the younger *Ss*, they were not spontaneously emitted when the child described the scenes.

This result raises a question about the role of verbal processes in the child's memory. Data of the younger children showed that verbal descriptions *per se* cannot explain why the unitized display is better remembered. In fact, as far as we could tell, the younger child showed no difference at all in describing the two kinds of displays.

Therefore, we cannot explain the advantage of Cond. U in terms of verbal processes. Better-remembered unitized scenes do not necessarily elicit different words from the nursery school child.

EXPERIMENT IV

Was verbalization of any importance in Exps. 1 and 2? Both experiments showed that unitized displays are better remembered, but both experiments had the subject *name* each object or picture when it first appeared. Was that part of the procedure necessary?

Several studies have shown the importance of labels on children's memory (e.g., Reese, 1962). If *S* names an item shortly before it is removed, he seems to remember it better; otherwise his performance is depressed. Now if Cond. U's advantage is related to the labeling process, then this advantage should disappear when *S* does not name each item. If Cond. U loses its superiority, then the naming of items is an essential part of the experimental procedure. Experiment IV therefore replicated Exp. II with one modification: *S* did not name each item.

This change in procedure posed a methodological problem. It is convenient to have children name each item in order to determine the child's name for that item. If an occasional child calls a cat a "bunny," then *E*

can subsequently expect "bunny" as the correct response. Therefore, we still wanted each child to name the object or picture, but not during the memory task.

Method

First, *S* was shown all 22 objects and all eight pictures in a randomized order. He was asked to name each item, and *E* recorded his response. Then the procedure of Exp. II was replicated. When each display appeared, though, *S* was not asked to name the ingredients.

There were 24 *S*s from the Bing Nursery School of Stanford University. Twelve *S*s (six males and six females) ranged from 3 years 6 months, to 4 years. The remaining 12 *S*s (six males and six females) ranged in age from 4 years 6 months, to 5 years. (The *S*s were comparable in age to the *S*s of Exp. II.)

Results and Discussion

Table 6 reports the mean number of correct responses for each experimental condition. In general, the means of Table 6 are low. In fact, every mean of Table 6 is lower than the corresponding mean of Table 3.

TABLE 6
MEAN NUMBER OF CORRECT RESPONSES

	Cond. U	Cond. S
Pictures (Max. score = 2)		
Age 3 1/2–4	1.08	1.16
Age 4 1/2–5	1.58	1.16
Objects (Max. score = 3)		
Age 3 1/2–4	2.25	1.42
Age 4 1/2–5	2.66	2.00

First consider the data for object displays. In these data, Cond. U was superior to Cond. S. An analysis of variance showed that the difference between the experimental conditions was significant; $F\ (1,20) = 13.78$, $p < .01$. No other variable or interaction reached statistical significance, except the age-difference, which was marginally significant: $F\ (1,20) = 3.42$, $p < .10$. Thus, the difference between Cond. U and Cond. S was still present, though performance was generally lower than that of Exp. II. On the other hand, the data for picture displays did not show significant differences. An analysis of variance was performed on these data, and all F values < 2.60, $p > .10$.

Picture displays occurred earlier in the testing than object displays. Therefore, the results of this experiment may reflect a difference between pictures and objects, or it may reflect a difference between early testings and later ones. Perhaps even without labeling, *S* can better remember a unitized display if he has had enough practice with the memory task.

Therefore, 10 more *S*s were tested, with the order of testing reversed. Object displays were tested first and picture displays afterwards. The results were very similar to those of Table 6. The mean number of objects recalled was 2.6 in Cond. U and 1.8 in Cond. S. This difference was statistically significant: $t\ (9) = 3.21$, $p < .02$. On the other hand, the recall of pictures showed no significant difference; the mean number of pictures recalled was 1.7 in Cond. U and 1.5 in Cond. S; $t\ (9) = 0.61$, $p > .10$. This pattern of results was just like that described above.

Thus, Cond. S and Cond. U did not differ significantly for picture displays unless *S* named each picture as it appeared. The naming procedure may make the child pay closer attention to the picture, or "orient him" better towards it. Pictures, after all, are more symbolic than objects; without some orienting task, *S* may not always perceive the stimulus. Or, if he does, he may fail to engage in whatever cognitive activity safeguards it from forgetting. At least with pictures, then, naming seems to be an important part of the procedure of Exps. I and II.

REFERENCES

Ekstrand, B. R. Backward associations. *Psychological Bulletin,* 1966, **65,** 50–64.

Hagen, J. W., and Kingsley, P. R. Labeling effects in short-term memory. *Child Development,* 1968, **39,** 113–121.

Hollingworth, H. L. *Psychology: its facts and principles.* New York: Appleton, 1928.

Horowitz, L., Norman, S. A., and Day, R. S. Availability and associative symmetry. *Psychological Review,* 1966, **73,** 1–15.

Horowitz, L. M., and Prytulak, L. S. Redintegrative memory. *Psychological Review,* 1969, **76** (in press).

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton, 1953.

Marchbanks, G., and Levin, H. Cues by which children recognize words. *Journal of Educational Psychology,* 1965, **56,** 57–61.

Reese, H. W. Verbal mediation as a function of age level. *Psychological Bulletin,* 1962, **59,** 502–509.

Reese, H. W. Imagery in paired-associate learning in children. *Journal of Experimental Child Psychology,* 1965, **2,** 290–296.

Sutcliffe, J. P. A general method of analysis of frequency data for multiple classification designs. *Psychological Bulletin,* 1957, **54,** 134–137.

Suzuki, N. S., and Rohwer, W. D. Verbal facilitation of paired-associated learning: Type of grammatical unit vs. connective form class. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7,** 584–588.

TWEDT, H. M., AND UNDERWOOD, B. J. Mixed vs. unmixed lists in transfer studies. *Journal of Experimental Psychology,* 1959, **58,** 111–116.

WILSON, K. V. A distribution-free test of analysis of variance hypotheses. *Psychological Bulletin,* 1956, **53,** 96–101.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 389–401 (1969)

Some Effects of Novelty and Overtraining on the Reversal Learning of Retardates[1]

FRANK D. TURRISI[2] AND BRYAN E. SHEPP
Brown University

Three experiments were conducted on retardates to assess the effects of the introduction of novel stimuli along an irrelevant dimension during a reversal, of different amounts of training, and of delay of reversal. It was found that the introduction of novel irrelevant stimuli (1) interrupts first-trial perseverative choice responding when *S*s have received minimal amounts of training but not when a strong criterion or overtraining is used, and (2) increases the latencies of responses to other stimuli along the relevant dimension. The ORE was found when a 24-hour period intervened between training and reversal but not when reversal immediately followed training.

The attention theory of discriminative learning of Zeaman and House (1963) has been successful in describing the data of various types of reversal and nonreversal shifts in retarded children (e.g., House and Zeaman, 1962; Campione, Hyman, and Zeaman, 1965) and rats (Shepp and Eimas, 1964). This theory, in common with other "chaining" theories (Goodwin and Lawrence, 1955; Sutherland, 1959; Lovejoy, 1966) postulates the acquisition of an attentional, or observing, response which precedes the learning of an instrumental response in a discriminative task. Specifically, Zeaman and House (1963) proposed that an attentional response is acquired to a relevant dimension, i.e., a dimension (e.g., color) along which stimuli (e.g., red and blue) are differentially correlated with reinforcement. As the attentional response to the relevant dimension is being learned, an instrumental response to the positive stimulus is also learned. The probability of attending to the relevant dimension (Po_1) is assumed to increase both as a function of reinforced instrumental responses to cues along that dimension and as a function of nonreinforced responses to cues along an irrelevant dimension (i.e., a dimension along which no stimuli are differentially correlated with reinforcement). The

[1] This research was supported by PHS Grants HD-01349-01 and 1-Fl-MH-30, 613-01 (MTLH). The investigators express appreciation for the cooperation and assistance of Dr. John Smith, Superintendent of Ladd School.

[2] Now at Florida State University.

probability of an instrumental response (Pr) to a cue along a given dimension, once that dimension has been attended to, increases as a function of the number of reinforced responses to that cue and of nonreinforced responses to any other available cues. Conversely, extinction of both responses (attentional and instrumental) occurs as a function of nonreinforced instrumental responses to a cue along the observed dimension. Extinction of an attentional response occurs indirectly when observing responses to other dimensions are followed by reinforced instrumental responses.

The probability of attending to a dimension may be affected by operations other than reinforcement. Zeaman, House, and Orlando (1958) found evidence which led them to suggest that the introduction of one novel stimulus along the relevant dimension makes possible the learning of a previously unsolved problem. Moreover, the novel stimulus was effective when replacing either the positive or the negative stimulus, indicating that the effect was not the result of a "simple preference for or aversion to novel or familiar stimuli." From this, it was concluded that "novelty enhances $Po_{(1,0)}$" where $Po_{(1,0)}$ is the initial value of Po_1 (Zeaman and House, 1963, p. 199).

A further generalization which might be made is that novelty, defined as the introduction of previously absent stimuli, enhances $Po_{2,3,\ldots,n}$. Stated explicitly, the introduction of novel stimuli along an irrelevant dimension raises the probability of attending to that dimension. Since Po is a theoretical construct and not amenable to direct empirical test, it is necessary to deduce, from the constructs of the theory, hypotheses which are empirically testable.

Hypothesis I. Consider the case in which *S*s are trained on a problem with one relevant dimension and one or more irrelevant dimensions. After such training, Po_1 should be high and Po_2, Po_3, ..., Po_n, the probabilities of attending to the irrelevant dimensions, should be low. Moreover, the value of Po_1 increases directly as a function of the number of trials on which reinforcement follows correct instrumental responses. It may then be hypothesized that novel stimuli along an irrelevant dimension would affect performance less if introduced following overtraining than if introduced at some prior point during the acquisition of a discriminative response.

One test of the preceding hypothesis would be to train *S*s on a two-choice visual discrimination and reverse them in a 2×2 factorial design varying level of training (overtraining vs. no overtraining) and irrelevant stimuli (novel stimuli vs. training stimuli). In the reversal paradigm, the attentional response which was relevant in the training problem

remains appropriate in the shift problem, but the reinforcement correlations of the stimuli are reversed. The introduction of novel stimuli along an irrelevant dimension might be expected to interfere with reversal learning to the extent that these operations increase Po of that dimension. Furthermore, the increase in Po should decrease with increased amounts of overtraining. Thus, it was predicted that overtraining and novelty would interact in their effects on performance in a reversal problem such that there would be less difference between the effects of old and novel stimuli with increasing amounts of overtraining.

Hypothesis II. Several experiments (see Mackintosh, 1965, for a recent review) have found that reversals in rats are facilitated with increased overtraining on the training problem. Such a finding is easily deduced from the theory of Zeaman and House (1963). Everything else being equal, reversal learning is facilitated to the extent that the relevant attentional response is transferred from training to reversal. A high Po_1 results in greater transfer, and overtraining increases Po_1. It would then be predicted that overtrained *Ss* would reverse faster than nonovertrained *Ss*.

EXPERIMENT I

Method

Subjects

The *Ss* were 40 retardates, institutionalized at the Ladd School, North Kingstown, Rhode Island, with a mean MA of 80.3 months (range = 39–155.5 months), a mean IQ of 49.0 (range = 30–72) and a mean CA of 177.9 months (range = 111–216 months). None had been run previously in discrimination learning experiments.

Apparatus

The apparatus was a modified Wisconsin General Test Apparatus. A 30 × 12-inch tray, mounted on a 36 × 36-inch table, could be pushed under a one-way vision screen to present the stimuli and withdrawn to bait the foodwells. The two foodwells were 3 inches in diameter and ½ inch deep, centered in the tray and 12 inches apart on centers. The entire apparatus was painted a flat gray.

The stimuli were seven geometric forms (cross, circle, T, triangle, square, star, and diamond) cut from 3-inch squares of ⅝-inch pine and painted one of seven colors (red, black, blue, yellow, brown, orange, and green), yielding 49 different stimuli. Each stimulus was glued upright upon a 3½-inch square of gray plywood.

Procedure

General. The stimuli were randomly assigned to *S*s, with the constraint that the cross and T were never assigned to any given *S*, either on the same problem or on consecutive problems. One-half of each experimental group was trained with color relevant and the other half with form relevant. For a given *S*, responses to one of the relevant stimuli (positive) were reinforced on all trials and responses to the other one (negative) on none of the trials. The stimuli from the irrelevant dimensions were paired with the relevant stimuli by a Gellermann series, and responses to them were reinforced on a 50% irregular schedule. For example, if form were relevant and color irrelevant, then two consecutive trials might be: (1) red circle (positive) vs. blue triangle (negative) and (2) red triangle (negative) vs. blue circle (positive).

Pretraining. The *S*s were verbally instructed to find the candy. Next the candy was presented in an uncovered foodwell for four trials with position randomized and equated. Finally, the candy was presented for three trials in each foodwell in random order, the foodwell containing the candy being covered with a $4 \times 5 \times \frac{1}{2}$-inch black plywood rectangle.

Experimental conditions. All *S*s were given 30 trials per day with an intertrial interval of approximately 12 seconds. Reinforcement consisted of one M & M candy on each trial. Following an initial choice, *E* said, "Good" if the response was correct or "No" if incorrect. Nothing was said following a correction response.

All *S*s were trained on a two-choice visual discrimination learning task (Problem I) to a criterion of nine correct of any 10 consecutive trials on a given day. The failure criterion was 150 trials. On the day following criterion day, Groups OTO and OTN were given 150 trials (5 days) of overtraining and then shifted to a new problem (Problem II). For Group OTO, the stimuli were the same as for Problem I, but the reinforcement contingencies of the relevant stimuli were reversed. For Group OTN, the reinforcement contingencies of the relevant stimuli were reversed and the irrelevant stimuli of Problem I were replaced by two different ones. Groups CO and CN were treated the same as Groups OTO and OTN, respectively, except that Problem II started on the day following criterion day with no interpolated overtraining. Criterion for Problem II was the same as for Problem I for all groups.

Results

In this and all subsequent experiments to be reported, trials to criterion yielded the same effects as errors to criterion; only the latter will be reported.

An analysis of variance of the mean number of errors through criterion (error scores) on the training problem revealed no significant main effects or interactions ($p > .05$).

An analysis of variance was performed on the mean error scores for the reversal problem and showed no significant main effects or interactions ($p > .05$).

Savings scores were calculated for each *S* by subtracting the error scores for Problem II from those of Problem I. Again there were no significant main effects or interactions ($p > .05$).

The frequencies of *S*s who perseverated in their responding to the positive training stimulus on the first trial of reversal were cast into a 2×4 contingency table. These data include three *S*s who did not complete the reversal problem but for whom first-trial data were available. Since more than 20% of the expected cell frequencies were less than five, the data were pooled in two ways for analysis (Siegal, 1956, pp. 178–179). Since Group CN was expected to be most affected by the introduction of novel irrelevant stimuli (Hypothesis I), it was tested against Groups CO, OTO, and OTN pooled. Fewer *S*s in Group CN than in the other three groups responded to the positive training stimulus on the first trial of the reversal ($\chi^2 = 7.79$, $df = 1$, $p < .005$). The effect of overtraining was tested by pooling the data for Groups CO and CN, and testing against the pooled data for Groups OTO and OTN. Groups CO and CN differed from Groups OTO and OTN ($\chi^2 = 4.95$, $df = 1$, $p < .03$) such that 18/21 *S*s in the overtrained groups perseverated whereas 12/22 *S*s in the nonovertrained groups perseverated.

Discussion

The lack of significant effects in both error scores and savings scores might suggest that neither novelty nor overtraining had an effect. On the other hand, the first-trial analysis indicates that the introduction of novel stimuli along the irrelevant dimension does in fact affect performance, although the effect on choice measures is transitory. It is important to note, however, that measures of performance on the first trial of reversal are reflecting changes in Po_1 which result primarily from the effects of stimulus novelty. On subsequent trials, the effects of the changed reinforcement contingency become manifest, superimposed upon, and possibly masking, any novelty effects which may obtain. Thus, weak novelty effects might be expected to occur on the first reversal trial but not over the whole course of the reversal.

The lack of an overlearning reversal effect (ORE) in the present experiment might be the result of the weak criterion used. Inspection of the data of Groups OTO and OTN revealed that 55% of the *S*s returned

to chance performance on the first day of overtraining, remaining there for 2 or more days before again achieving criterion. For these *Ss*, the criterion had been achieved fortuitously, and it is reasonable to expect a comparable effect for the nonovertrained *Ss* as well.

The effects of a weak criterion are at least twofold. First, *Ss* who do not start performing consistently above chance until the third or fourth day of overtraining have not received as much overtraining as *Ss* who have been at criterion throughout. The result of a weak criterion in Problem I would be to reduce the difference in amount of overtraining which would be produced between the overtrained and nonovertrained groups, and hence to attenuate the ORE. Second, if the weak criterion results in fast learning on Problem II, any differential transfer effects might be obscured by a floor effect. It is relevant to note that 31 of the 40 *Ss* in this experiment learned Problem II in 10 or fewer errors.

A second experiment was performed in an attempt to determine whether the initial hypotheses could be confirmed, with measures which reflect overall learning rate, by utilizing a stronger criterion of learning in both problems.

EXPERIMENT II

Method

Subjects

The *Ss* were 56 institutionalized retardates, with a mean MA of 78 months (range = 31–179 months), a mean IQ of 43 (range = 19–83), and a mean CA of 236 months (range = 130–352 months). All *Ss* were experimentally naive.

Apparatus

The apparatus and stimuli were the same as those used in Exp. I.

Procedure

With one exception, the same procedure was used here as was used in Exp. I. Instead of a criterion of nine correct of any consecutive 10 trials, *Ss* in Exp. II were trained to a criterion of 25 correct responses in a given session of 30 trials.

Results

An analysis of variance on error scores for Problem I showed no significant main effects or interactions ($p > .05$). Similarly, there were no significant main effects or interactions obtained for the error scores

for Problem II. The savings scores revealed a significant main effect for overtraining ($F = 8.59$, $df = 1/51$, $p < .01$). Overtrained *Ss* learned Problem II faster than they had learned Problem I whereas nonovertrained *Ss* learned Problem I faster than Problem II.

Regarding the first-trial data for Problem II, the groups are virtually identical, with responding significantly different from chance ($z = 5.21$, $p < .001$) in the direction of perseveration to the positive training stimulus. Three of the *Ss* in Groups OTO and OTN lost criterion during overtraining, and all of them regained it and kept it on the second day of overtraining.

Discussion

The purpose of Exp. II was to test the two original hypotheses under a stronger criterion of learning than that which was used in Exp. I. That the present criterion was in fact more rigorous is evidenced by the fact that, whereas 55% of the *Ss* in Exp. I had reverted to chance performance for 2 or more days during overtraining, only 10.7% of the *Ss* in this experiment did so and for only 1 day.

There are two major differences in the findings between Exps. I and II. First, an ORE was found in the present experiment. The presence of the ORE suggests that the explanations advanced for its absence in Exp. I were valid. With the stronger criterion only a small proportion of the *Ss* lost criterion during overtraining, insuring that all OTO and OTN *Ss* were in fact overtrained. In addition, only 18/54 of the *Ss* had error scores of 10 or less, as compared with 31/40 in Exp. I, making a floor effect much less probable.

The second major difference in the results of the two experiments is in the frequency of perseverative responses on the first trial of Problem II. It is possible that the stronger criterion used in the present experiment was sufficient to raise Po_1 to the point where the novel irrelevant stimuli were not attended to, in which case no disruption of perseverative responses would be observed.

Another possibility is that the novel stimuli along the irrelevant dimension were in fact attended to, but that the instrumental response was nevertheless made to stimuli along the relevant dimension. In the terms of attention theory, both the relevant and irrelevant dimensions were being attended to, but for different reasons: the relevant dimension because attentional responses to it had been consistently followed by reinforced instrumental responses, the irrelevant dimension because of the introduction of novel stimuli along it.

The above hypothesis is not testable using only choice measures in the present situation, since systematic responses to a given stimulus permit

inferences only about the dimension along which the stimulus lies. Hence, perseverative responding on the first reversal trial permits one to infer only that the relevant dimension was observed. One resolution of this ambiguity would be the use of a measure which may vary independently of choice measures. One such measure is latency of response. If changes in response latency were found to be correlated with changes in stimuli along the irrelevant dimension when *S* is consistently choosing a stimulus along the relevant dimension, then one might reasonably infer that both dimensions were being attended to. This, in spite of that fact that responding to one dimension only is being reflected in the choice measures.

To test this notion, a third experiment was run. Experiment III is a replication of Exp. II, except that *S*s were reversed immediately following acquisition or overtraining in an attempt to insure that *S*s would be responding to the positive training stimulus on the first trial of reversal. In addition to the usual choice measures, response latencies were taken.

EXPERIMENT III

Method

Subjects

The *S*s were 38 institutionalized retardates, with a mean MA of 80 months (range = 32–156 months), a mean IQ of 46 (range = 27–87), and a mean CA of 248 months (range = 129–343 months). All were experimentally naive.

Apparatus

The same apparatus was used as had been used in Exps. I and II, except that a Standard Timer was added to measure response latencies. The timer was started by a microswitch when the tray was in a position such that *S* could see the stimuli, and was stopped when *S* dislodged either stimulus object from the foodwell.

Procedure

This experiment replicated Exp. II except for the length of delay between the end of the training or overtraining and the start of reversal learning. On the first day of Problem II, *S*s were given 10 trials of Problem I before reversal. The first trial of Problem II was presented with no perceptible change in intertrial interval from the last trial of Problem I. It was required that eight or more of the ten trials of Problem I be

correct before starting Problem II, a criterion which all *Ss* met on the first attempt.

Results

An analysis of variance of the error scores for Problem I showed no reliable main effects or interactions. Analysis of the error scores and savings scores for Problem II likewise yielded no reliable effects ($p > .05$).

All *Ss* responded to the positive training stimulus on the first trial of the reversal, a highly significant ($z = 5.91$, $p < .001$) departure from chance.

The mean latencies are plotted (See Fig. 1) in blocks of five trials for the last ten trials of Problem I and the first 30 trials of Problem II. The mean response latencies on the first trial of the reversal are also shown in Fig. 1. An analysis of variance was performed on the latencies of the last 10 trials of Problem I, and no differences were found, indicating that all *Ss* were performing at the same level prior to reversal. Difference scores (D_1) were obtained for each *S* by subtracting the latency of the first trial of reversal from the mean of the immediately preceding five trials of Problem I. The scores reflect systematic changes in response

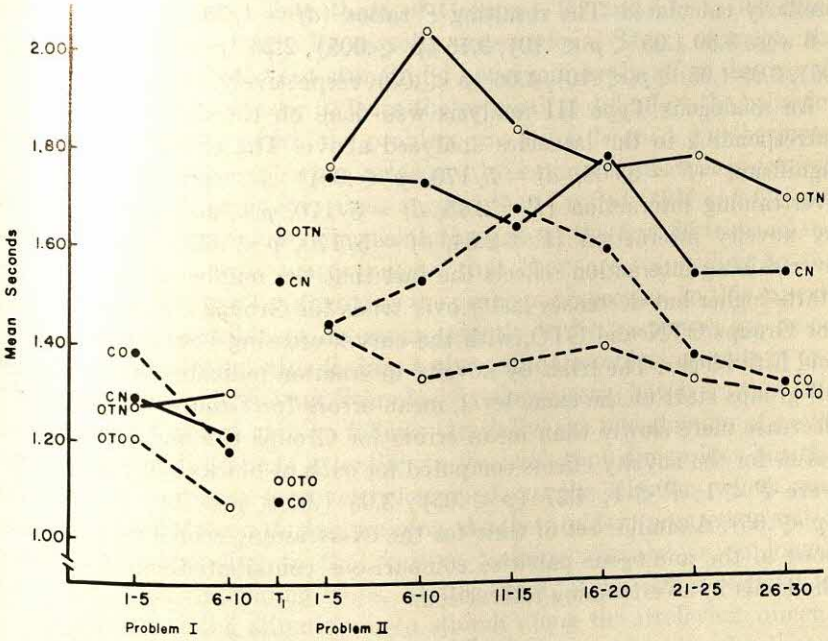


FIG. 1. Mean response latencies for the last 10 trials of Problem I, the first trial of Problem II (T_1), and the first 30 trials of Problem II.

latency which result solely from the introduction of novel irrelevant stimuli. A further set of difference scores (D_2) was derived for each *S* by subtracting the mean of the first five trials of Problem II from the immediately preceding five trials of Problem I. These scores presumably reflect both novelty effects and the effects of the reversed reinforcement contingency. An analysis of variance of the D_1 scores revealed a significant novelty effect ($F = 16.21$, $df = 1/34$, $p < .001$) but no other main effect or interaction. An analysis of the D_2 scores showed no significant main effects or interactions. Since half of the *S*s learned the reversal in the first 30 trials, the data for those trials were subjected to a finer analysis by examining the mean latencies and the mean errors in six blocks of five trials each.

A Lindquist Type III analysis of variance was performed on the blocked latency scores for the first 30 trials of reversal (Fig. 1). The latencies for Groups OTN and CN were significantly ($F = 5.12$, $df = 1/34$, $p < .05$) longer than those for Groups OTO and CO. The trials effect was also significant ($F = 2.98$, $df = 5/170$, $p < .05$), with latencies decreasing across trials. None of the other main effects or interactions were significant. The mean squares for the novelty effects computed for each block of trials were tested against error mean squares which were similarly calculated. The resulting *F* ratios ($df = 1/34$) for trial blocks 1–6 were 3.30 ($.05 < p < .10$), 9.16 ($p < .005$), 2.26 ($p > .05$), 2.46 ($p > .05$), 3.95 ($.05 < p < .10$), 4.66 ($p < .05$), respectively.

An analogous Type III analysis was done on the mean error scores corresponding to the latencies analyzed above. The effect of trials was significant ($F = 55.08$, $df = 5/170$, $p < .001$) as were the trials by overtraining interaction ($F = 2.45$, $df = 5/170$, $p < .05$) and the trials by novelty interaction ($F = 2.54$, $df = 5/170$, $p < .05$). The trials by overtraining interaction reflects the fact that the number of mean errors starts higher but decreases faster over trials for Groups CN and CO than for Groups OTN and OTO, with the curves crossing between the fourth and fifth blocks. The trials by novelty interaction indicates that, although all groups start at the same level, mean errors for Groups CN and OTN decrease more slowly than mean errors for Groups CO and OTO. The *F* ratios for the novelty effects computed for each of blocks 1–6 ($df = 1/34$) were $F < 1$; $F < 1$; 4.27 ($p < .05$); 3.05 ($.05 < p < .1$); $F < 1$; 5.10 ($p < .05$). A similar set of tests for the overtraining scores revealed that none of the analogous pairwise comparisons contributed significantly to the trials by overtraining interaction.

Discussion

In this experiment, neither Hypothesis I nor Hypothesis II was confirmed with overall error scores, with savings scores or with first-trial

choice scores. Although novelty did not interact with overtraining in the predicted manner, it did affect behavior in Problem II. It was found that *Ss* who had been extensively trained on a dimension, and who were responding systematically to stimuli along that dimension, could nevertheless attend to another dimension when novel stimuli were introduced along it. Although all *Ss* made the same choice response on the first trial of reversal, the latencies of Groups CN and OTN increased relative to the pre-reversal level, whereas those of Groups CO and OTO did not. Since, from the point of view of *S*, nothing had changed prior to the response except the irrelevant stimuli for Groups CN and OTN, it is safe to infer that the first-trial effect resulted from that change. While choice and latency measures seem to be varying independently on the first trial of reversal, the results for subsequent trials are not so clearcut. With respect to latencies, the significant novelty effect and the significant trials effect taken together with the lack of a trials by novelty interaction indicate that Groups OTN and CN were taking longer to respond than Groups OTO and CO throughout the first 30 trials of reversal, but that the latencies for all four groups were decreasing at the same rate as the problem was learned. The choice measures reflect a quite different sort of behavior, since the error rates of Groups OTN and CN fell off at an increasingly lower rate than those of Groups OTO and CO as indicated by the significant trials by novelty interactions. Thus, although the addition of novel irrelevant stimuli did not significantly affect the overall number of trials necessary to learn the reversal, it did affect the behavior on the trials during which the problem was being learned. Furthermore, the effects were different when measured by latency than when measured by error scores. The latency data suggest that Groups OTN and CN are attending to the irrelevant dimension throughout the first 30 trials of the reversal and the error scores suggest that they are responding to the novel stimuli to a greater degree as the problem progresses. In this experiment, it is not possible to determine exactly which responses are being made to the irrelevant stimuli since a given stimulus object consists of one relevant and one irrelevant stimulus. It does seem, however, that the course of the reversal goes as follows: Initially, the introduction of novel irrelevant stimuli attracts attention to the irrelevant dimensions but the instrumental response is nevertheless made to the stimulus which was consistently reinforced during training. As the problem progresses and *S* learns that responding to the positive training stimulus is no longer reinforced, such responding stops. At this point, *S* can either respond to the negative training stimulus or to stimuli along the irrelevant dimension. The data of this experiment suggest that the latter predominates when the stimuli along the irrelevant dimension are novel, but that the former predominates when no change is made in the irrelevant dimension.

It is likely that both dimensions are responded to, to a degree, but that the novelty groups respond predominantly to the irrelevant dimension whereas the others respond predominantly to the relevant dimension.

Conclusions

The present study fails to provide strong evidence in support of Hypothesis I. Under no conditions was overtraining found to interact with novelty, on either learning or latency measures. At best, it can be stated that the introduction of novel irrelevant stimuli interrupts first-trial perseverative choice responses in reversals when *Ss* have received minimal amounts of training, but does not when a strong criterion or overtraining is used.

Qualified support was found for Hypothesis II. The ORE was found in the retarded children of this study when a 24-hour period intervened between training and reversal, but not when reversal followed immediately after training. While reversals have been found to be facilitated by delay, a similar finding for the ORE is unique, and no explanation for it is immediately obvious within the boundary of current theory.

The finding that the choice responses behaved differently from the latencies in Exp. III merits a comment. The difference may reflect different processes or may simply reflect differences in the sensitivities of the measures. In either case, since latency data are virtually nonexistent in the discrimination literature for human *Ss*, the problem would seem to warrant more careful scrutiny in the future.

REFERENCES

Campione, J., Hyman, L., and Zeaman, D. Dimensional shifts and reversals in retardate discrimination learning. *Journal of Experimental Child Psychology,* 1965, **2,** 255–263.

Goodwin, W. R., and Lawrence, D. H. The functional independence of two discrimination habits associated with a constant stimulus situation. *Journal of Comparative and Physiological Psychology,* 1955, **48,** 437–443.

House, B. J., and Zeaman, D. Reversal and nonreversal shifts in discrimination learning in retardates. *Journal of Experimental Psychology,* 1962, **63,** 444–451.

Lovejoy, E. Analysis of the overlearning reversal effect. *Psychological Review,* 1966, **73,** 87–103.

Mackintosh, N. J. Selective attention in animal discrimination learning. *Psychological Bulletin,* 1965, **64,** 124–150.

Shepp, B. E., and Eimas, P. D. Intradimensional and extradimensional shifts in the rat. *Journal of Comparative and Physiological Psychology,* 1964, **57,** 357–361.

Siegal, S. *Nonparametric statistics for the behavioral sciences,* New York: McGraw-Hill, 1956.

Sutherland, N. S. Stimulus analysing mechanisms. In *Mechanization of thought processes,* Teddington, England: National Physical Laboratory Symposium No. 10, 1959, Vol. 2, pp. 575–609. London: Her Majesty's Stationery Office.

ZEAMAN, D., AND HOUSE, B. J. The role of attention in retardate discrimination learning. In Ellis, N. R. (Ed.), *Handbook of mental deficiency*, New York: McGraw-Hill, 1963.

ZEAMAN, D., HOUSE, B. J., AND ORLANDO, R. Use of special training conditions in visual discrimination learning with imbeciles. *American Journal of Mental Deficiency*, 1958, **63**, 453–459.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 402–409 (1969)

Young Children's Initial and Changed Names for Form-Color Stimuli[1]

REGINA A. MODRESKI[2] AND ALBERT E. GOSS
Douglass College, Rutgers—The State University of New Jersey

In naming form and color stimuli, 3-, 4-, and 5-year-old children responded initially more often with conventional or adequately descriptive names for form than for color. Frequency of form names increased with age; frequency of color names decreased slightly. With few exceptions, naming was by form or color alone, not by both. Those children who were subsequently trained to name stimuli by form and color used both names more often than untrained children. Training increased frequency of color names more than that of form names.

This experiment investigated young children's naming responses to stimuli that differ in form and color. The objectives were: (a) to assess initial naming responses to these stimuli by males and females of ages 3, 4, and 5 years; (b) to determine effects of training with conventional names for form and color on initially incorrect and incomplete naming responses; and (c) to determine maintenance of and generalization of naming responses after training.

Data on initial naming responses complement data on children's preferences for form and color by motor choice (e.g., Brian and Goodenough, 1929; Kagan and Lemkin, 1961; Suchman and Trabasso, 1966). Such data and information on training and on maintenance and generalization are requisite to effective planning and theoretical interpretation of results of experiments on verbal mediation in children's conceptual behavior.

METHOD

Stimuli

For initial assessment and training, the stimuli were 12 form-color combinations. Red, blue, green, and yellow construction paper was cut into squares, circles, and equilateral triangles, each 1 square inch in area. Each figure was centered on a white, 5 × 5-inch card.

[1] R. A. Modreski undertook this research as an NIMH fellow at Douglass College during the summer of 1967. It was completed with support through Public Health Service Research Grant MH 13531-01.

[2] Now at the Department of Psychology, Pennsylvania State University.

Added for the posttraining assessment were rectangles, ellipses, and isosceles triangles, also 1 square inch in area, in each of the four colors. Twelve figures had the long axis horizontal and 12 other had the long axis vertical.

Procedure

Pretraining assessment. Pretraining assessment was the same for both experimental and control conditions. Each of the 12 stimuli was presented once, and then again, for about 10 seconds each time. The stimuli were in a different random order for each trial, and each *S* received a different order of presentation. The children were tested individually under instructions to "Tell me whatever you can about these pictures." No informative feedback about responses was given.

Training. The first of two successive, daily, training sessions began immediately after the initial assessment. Children of the experimental condition had instructions and conditions like those of a paired-associates task with an anticipation format. The form-color combinations were stimulus members. Conventional names for the color and form of each stimulus, spoken by *E* in that order, were response members. Approximately 5 seconds were allowed for anticipation of color and form names.

First-session training was to a 22/24 criterion or for five more trials. Second-session training the next day was to the 22/24 criterion or for five more trials. Thus, total training was to a repeated criterion or 10 trials.

Children of the control condition continued to see and name the stimuli for three additional trials, or until they became restless. No information was provided by *E*. They returned the next day for five more trials. Thus, these *S*s had up to eight trials.

Posttraining. Immediately after the last training trial, the 36 stimuli of the enlarged set were each presented once. A different random order of presentation was used for each *S*. Instructions and conditions were those used in pretraining.

Subjects

Five males and five females, ages 3, 4, and 5 years, were assigned in order of availability to a counterbalanced, prescheduled experimental or control condition. The respective age ranges in months were 36–46, 49–59, and 61–71. The mean ages for males and females in each condition were approximately equal within an age level; the respective means were 41.25, 53.75, and 64.25 months.

The *S*s were obtained primarily by mail solicitation of parents who had children of the desired ages and permitted them to participate. Parents were informed of the purpose and nature of the study. The solicitation

and returns were primarily among parents, one or both of whom were graduate students at Rutgers University in New Brunswick, and who were living in University housing. The *Ss* were predominantly native-born whites; one was a native-born black, and two were from a Puerto Rican, white family. Two *Ss* were of Chinese origin, and one each was of Egyptian, Indian, Israeli, and Italian origin. All of the latter *Ss* had been in the United States at least a year. All spoke understandable English, and named in a way similar to the native-born *Ss* of comparable age and sex. No more than one foreign-born *S* was in any combination of age, sex, and conditions.

RESULTS

Pretraining assessment. Names for particular stimuli were categorized as conventional, adequately descriptive (e.g., "block" for square, "ball" for circle, "tent" for triangle), and inadequately descriptive (e.g., "black" for green, "sun" for green square). Responses in the first two categories which corresponded with the particular form or color of a stimulus were considered "correct." Responses that did not correspond (e.g., "blue" for green; "circle" for square) were considered "incorrect," as were inadequately descriptive names and failures to name one or both values along a dimension. The range of correct responses for each attribute was 0–12 per trial.

Table 1 contains confusion matrices for correct and incorrect conventional color and form responses to each stimulus. Frequencies at each age are for males and females of experimental and control conditions during both trials. Correct and incorrect adequately descriptive names, and inadequately descriptive names are markedly variable, and particular names in these classes occur infrequently. Therefore, these responses are not included in the confusion matrices.

For each color of each form, frequencies of correct form names increase and frequencies of incorrect form names decrease with age. For each form, frequencies of correct use of "red" decrease with age. Frequencies of correct use of "blue," "green," and "yellow" are relatively constant across the three ages. Square stimuli and red stimuli are named correctly somewhat more often than the other forms and colors.

Frequencies of both correct conventional and correct adequately descriptive names on each trial were also determined for each *S*. Inclusion of the latter responses provides more complete representation of *Ss*' naming by form and by color. Table 2 shows means of the totals of both classes of responses during Trial 2 for each combination of age, sex, and condition. Trial 2 is used because training began after this trial, and its over-all mean is slightly higher than that for Trial 1. However, the dif-

TABLE 1

CONFUSION MATRICES OF FREQUENCIES OF CONVENTIONAL FORM NAMES (Square, S; Circle, C; Triangle, T) AND OF CONVENTIONAL COLOR NAMES (Red, R; Blue, B; Green, G; Yellow, Y) FOR EACH FORM-COLOR STIMULUS BY *Ss* OF AGE 3, 4, AND 5 YEARS[a]

		3	4	5	3	4	5	3	4	5	3	4	5
Form	Color	S			C			T					
S	R	10	15	24	1	1		1		1			
	B	10	15	26	1	1		2		1			
	G	13	16	26	2	2							
	Y	10	16	26	5			1					
C	All	43	62	102	9	4		4		2			
	R				4	11	19						
	B				4	12	17						
T	G	1			5	10	17						
	Y				1	9	18						
	All	1			14	42	71						
	R	9		2				4	13	22			
	B	9	4	3				4	10	22			
	G	8	4		1			5	11	20			
	Y	7	1	4				7	12	21			
	All	33	9	9	1			20	46	85			
		R			B			G			Y		
S	R	13	9	10		1							
C		12	8	9	1								
T		13	10	7				1					
All		38	27	26	1	1		1					
S	B	1			7	6	9				1		
C		1	1		8	7	9	1					
T		3			7	8	8	2					
All		5	1		22	21	26	3			1		
S	G	1						11	9	9			
C		1						9	9	11			
T		2						9	9	5			
All		4						29	27	25			
S	Y	2									8	12	8
C		1						1			8	10	10
T		2									8	11	7
All		5						1			24	33	25

[a] Row frequencies for each stimulus at each age are based on 40 possible responses by 20 *Ss* during two trials.

TABLE 2

MEANS OF TOTALS OF CORRECT CONVENTIONAL AND ADEQUATELY DESCRIPTIVE NAMES FOR FORM (F) AND COLOR (C) VALUES OF TRAINING STIMULI FOR MALE (M) AND FEMALE (F) *Ss* OF EXPERIMENTAL (E) AND CONTROL (C) CONDITIONS DURING TRIAL 2 OF PRETRAINING AND DURING POSTTRAINING

			Pretraining		Posttraining	
Age	Sex	Condition	F	C	F	C
3	M	E	6.6	2.6	8.6	8.6
		C	1.6	2.4	4.0	3.0
	F	E	5.8	4.2	6.8	7.0
		C	3.8	2.8	4.0	3.2
4	M	E	6.8	5.6	11.6	11.8
		C	5.6	4.6	6.6	5.2
	F	E	8.6	0.0	9.8	4.6
		C	8.8	0.8	9.6	0.2
5	M	E	11.0	0.0	10.4	8.4
		C	7.2	2.7	10.0	3.8
	F	E	8.4	4.6	9.4	6.8
		C	7.4	5.0	9.8	4.8

ference between these means was not significant ($F(1,48) = 3.31$, $p > .05$), and trials did not enter into any significant interaction.

Across combinations, form names occurred more often than color names ($F(1,48) = 20.35$, $p < .01$). The difference held for combinations separately, 3-year-old boys of the control condition excepted. Form names increased with age; color names decreased slightly. Across form and color, correct names increased with age ($F(2,48) = 6.18$, $p < .01$), but the interaction of age and form-color names fell short of significance ($F(2,48) = 2.71$, $p > .05$). Sex and, for initial assessment, the dummy variable of conditions made no difference either across other combinations or in any interaction.

Whether or not color names and form names for particular stimuli are correct, they may be in patterns of relationships to stimuli that constitute responding by color or by form separately or together. Frequencies of incorrect adequately and inadequately descriptive color names (e.g., "black" for green) and form names (e.g., "square" for circle) were added to those for correct names to yield frequencies of all color names and of all form names. For 3-year olds, frequency of all form names increased more than that of all color names. For 4- and 5-year olds, there were only slight increases in these frequencies. Although less pronounced, the increase in

frequencies of form names with age held, as did the slight decrease in color names.

Examination of individual protocols showed that most *Ss* named by form or by color alone. Only three of the 60 children named frequently and consistently by both form and color.

Training. Only 6 of 30 *Ss* in the experimental condition failed to reach 22/24 correct within 10 trials. Three were 3-year olds, one was a 4-year old, and two were 5-year olds. Trials to criterion decreased with age, but the overall differences were not significant ($F(1,24) = 2.50$, $p > .05$). A significant relationship held between trials to criterion and totals of correct names on Trials 2 of the initial assessment (Pearson $r(28) = -.60$, $p < .01$).

Posttraining. Names for the 12 stimuli of pretraining-training were analyzed separately from names for the 24 new stimuli. Table 2 also shows means of totals of correct color names and form names for the former stimuli. Across age and sex, and for the six combinations separately, training increased frequencies of color responses relative to frequencies during pretraining and to frequencies for the control condition. Training produced much smaller increases in frequencies of form names and, for 5-year-old males, there was a slight decrease. Those increases that occurred were in conventional names rather than in adequately descriptive names. For the control condition, most of the changes were slight increases.

The critical comparisons are between experimental and control conditions during posttraining. Form names occurred more often than color names ($F(1,48) = 10.31$, $p < .01$). Both form and color names occurred more often for the experimental than for the control condition ($F(1,48) = 25.42$, $p < .01$); for males than for females ($F(1,48) = 4.21$, $p < .05$); and with increasing age ($F(2,48) = 4.33$, $p < .05$). Of the interactions of these variables, only that for conditions and sex approached significance ($F(1,48) = 3.06$, $p > .05$). The difference between experimental and control conditions was greater for boys than for girls. The same patterns of differences obtained for all form and all color names.

For the additional stimuli, conventional and adequately descriptive form names ($\bar{X} = 13.03$), occurred more often than did color names ($\bar{X} = 9.75$, $F(1,48) = 3.78$, $p < .05$). Form and color names combined occurred more often for the experimental condition ($\bar{X} = 13.73$) than for the control condition ($\bar{X} = 9.05$, $F(1,48) = 11.83$, $p < .01$). The $F(2,48) = 3.38$, $p < .05$ for means of form and color names combined of 8.92, 12.25, and 13.00 for 3-, 4-, and 5-year olds, respectively. The pattern of the interaction of age, sex, and form-color names ($F(2,48) = 3.68$, $p < .05$) is neither described nor interpreted easily. Sex appeared in no other significant *F*s. Al-

though the difference between means of form names and color names was less for the experimental than for the control condition, the interaction was not significant.

The pattern of differences obtained for all names was similar to that for correct names. However, the *F*s for age and for the interaction of age, sex, and form-color names were not significant.

Examination of individual protocols of *S*s of the experimental condition showed that 21 and 15 *S*s, respectively, responded consistently to the 12 training and 24 new stimuli with names for form and color. Five and nine *S*s responded by form, two and four by color. For the control condition only three and four *S*s, respectively, responded to the training and new stimuli with names for form and color. Eighteen and 18 *S*s responded by form, eight and seven by color. The remaining *S*s of these conditions did not respond in discernible patterns.

DISCUSSION

Pretraining. Even 3-year-old children named by form more often than by color. The difference increased with 4- and 5-year-old children. Three-year-old children typically choose by color rather than by form in matching-to-sample preference tests (Brian and Goodenough, 1929; Suchman and Trabasso, 1966). By 4 years or slightly older, choice by form is more likely than choice by color, with some exceptions such as Brian and Goodenough's 4- and 5-year olds. Thus, the patterns of initial names and motor choices of 3-year-old children are seemingly incongruent while those of the names and choices of 4- and 5-year-old children are seemingly congruent. Congruence of patterns of names and choices is consistent with but not unequivocally supportive of Kagan and Lemkin's (1961) suggestion that implicit naming responses influence motor choice. However, the studies noted here, and other studies not mentioned explicitly, differ markedly with respect to kinds and numbers of specific stimuli, manners of their presentation, response requirements, details of characteristics of the children, and criteria for scoring or including responses. Therefore, agreements or disagreements between patterns of names and of motor choices of young children are at best tentative and of limited generality.

For both groups and individuals, young children's initial naming responses to form-color stimuli are not equiprobable. Indeed, they are likely to respond with form names alone, or with color names alone. In studies of conceptual behavior, particularly those that involve verbal mediation, initial naming responses could be determined. Subsequent verbal pretraining could be specific to particular initial patterns of relationships between stimuli and names.

Training. In relatively few trials conventional names for the colors and

forms were used more correctly and completely. The speed with which children acquired these names tended to increase with age.

Posttraining. During posttraining, the 12 training stimuli were presented with 24 new stimuli, and *E* no longer provided correct conventional names. The names acquired during training were maintained during the posttraining trial, although at frequencies below those of the last training trial. Also, these names generalized to the 24 new stimuli. This outcome could be described as training-induced "conservation" of form.

REFERENCES

BRIAN, C. R., AND GOODENOUGH, F. L. The relative potency of color and form perception at various ages. *Journal of Experimental Psychology,* 1929, **12,** 197–213.

KAGAN, J., AND LEMKIN, J. Form, color, and size in children's conceptual behavior. *Child Development,* 1961, **32,** 25–28.

SUCHMAN, R. G., AND TRABASSO, T. Color and form preference in young children. *Journal of Experimental Child Psychology,* 1966, **3,** 177–187.

Errata

Vol. 6, No. 4 (1968), in the article, "Looks by Preschoolers at the Experimenter in a Choice-of-Toys Game: Effects of Experimenter and Age of Child," by Lauren Harris, pp. 493–500:

Page 496, line 4 from top, " . . . mean proposition scores . . . " should read: " . . . mean proportion scores . . . "

Vol. 7, No. 2 (1969) in the article, "Dimension Preference and Performance on a Series of Concept Identification Tasks in Kindergarten, First-Grade, and Third-Grade Children," by Merrill M. Mitler and Lauren Harris, pp. 374–384:

Table 3 on p. 381 should appear as follows:

TABLE 3

MEAN NUMBER OF TRIALS TO CRITERION FOR FIRST DIMENSION (NUMBER) FOR CONTROL *Ss* AND COMPARABLE EXPERIMENTAL *Ss*

Level in school	Mean	
	Control *Ss*	Experimental *Ss*
Kindergarten	37.2 $N = 11$ (SD = 26.87)	57.8 $N = 9$ (SD = 24.55)
Third-grade	9.2 $N = 6$ (SD = 2.67)	18.8[a] $N = 4$ (SD = 16.4)
Total	$N = 17$	$N = 13$

[a] Because there were only four *Ss* in this experimental group-cell, four additional *Ss* were tested. A one-way analysis of variance still failed to disclose a difference between control and experimental third-graders on the number task. (Experimental mean = 23.9, $N = 8$, Control mean = 9.2, $N = 6$, $F = 1.48$, df, $df = 1/12$, $p > .40$.) It should be mentioned that the original mean of 18.8 and the new mean of 23.9 are distorted by two high scores and therefore are not representative.

7538—ACAD.—B—DEW—

Singer & Singer

PSYCHOLOGICAL DEVELOPMENT IN CHILDREN

Why do children develop as they do?

Here is a book which explains the development of children in terms of psychological processes rather than recording only development's outward manifestations. Designed for the basic undergraduate course in child psychology, this text integrates the chronological and theoretical approaches to form a unified view of psychological development from infancy to late adolescence. Such topics as stage theories, perception, language, and internalization are discussed within this theoretical framework. The authors give special attention to the cognitive development theories of Piaget, Bruner, Kagan and Kohlberg, and to social variables such as patterns of child rearing and learning problems of disadvantaged children.

Test questions are available.

By Robert D. Singer and Anne Singer, both of the University of California at Riverside.

About 450 pp. Illus. About $8.00.
Just Ready!

W. B. SAUNDERS COMPANY
West Washington Square
Philadelphia, Pa. 19105

Please send & bill me:

Singer & Singer: PSYCHOLOGICAL DEVELOPMENT IN CHILDREN
About $8.00

Name .....

Address ..... ZIP .....

JEC 10-69

ADJUSTMENT: MODELS AND MECHANISMS

By **Irving Tucker**
American University, Beirut

The author provides a comprehensive and systematic integration of those laboratory and clinical research findings that contribute to our current understanding of human personality and adjustment. While there is significant stress on the psychology of learning, the effects of physiological processes and social conditioning are also carefully examined. Numerous graphs and schematic sketches precisely illustrate points developed in the text.

CONTENTS:

The Behavior Science Approach. Biology and Stress. Motivation and Emotion. Principles of Learning—I. Principles of Learning—II. Frustration. Verbal Behavior and Thought Processes. Social Learning and Imitation. The Self Concept and the Learning of Anxiety. Repression-Sensitization and Mechanisms of Personality Defense. Complex Modes of Human Adjustment. Approaches to the Modification of Behavior.

In preparation

AP 2331

Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

Experiments in Second-Language Learning

By **Edward Crothers**
and **Patrick Suppes**
Both at the Institute for Mathematical Studies in the Social Sciences
Stanford University, Stanford, California

This volume views second-language learning in the light of mathematical learning theory. It describes a series of sixteen experiments involving the acquisition of Russian by native American college students and dealing with phoneme discrimination, phoneme-grapheme correspondence, and the learning of vocabulary, morphology and syntax. The authors formulate mathematical learning models and apply them to the experimental data.

The book is written as a series of technical experimental reports. Discussions of psycholinguistic factors appear separately from the application of models.

CONTENTS:

Some Remarks on Stimulus-Response Theories of Language Learning
Learning to Discriminate Russian Phonemes
Learning the Orthographic Representations of Russian Sounds
Vocabulary Learning
Morphology Learning
Grammar Learning
Suggestions for Future Research
REFERENCES.
AUTHOR INDEX—SUBJECT INDEX.

1967, 374 pp., $13.50

AP 2024

Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

A Volume in **SOCIAL PSYCHOLOGY**
A Series of Monographs, Treatises and Texts
Edited by **Leon Festinger** and **Stanley Schachter**

EXPERIMENTS IN SOCIAL PSYCHOLOGY

By **Paul G. Swingle**
McGill University

Stressing the development of research skills in experimental social psychology, this text exposes the student to a variety of research areas and problems of methodology including observer bias, brainstorming, conformity, psychological gaming, and attitude change.

Designed for use in experimentally oriented undergraduate courses in conjunction with a basic social psychology text, this volume provides stimulating experimental research projects requiring no specialized equipment. Statistical study is not a prerequisite, since the student is encouraged to plot, examine and understand his data rather than concern himself with establishing the reliability of differences.

A comprehensive introduction precedes each project, and reprinted articles provide theoretical background.

1969, 292 pp., $7.50

AP 2332

Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

INFORMATION FOR AUTHORS

The *Journal of Experimental Child Psychology* will publish papers in which the behavior and development of children is clearly related to its determining variables. Typically (not always) this would mean that a variable has been manipulated in an experimental manner. Articles will be reviewed in terms of the apparent reliability of their findings and the clarity of the relationships they affirm. Research conceived in any theoretical framework is invited, so long as its conclusions are reliable and compelling. Occasionally, critical reviews of related bodies of experimental literature, theoretical discussions which define new paths of experimental work, and well-controlled developmental studies on infra-human subjects will be published. The criterion for acceptance of such papers will be their value in prompting experimental research in related areas with children. Short notes describing new apparatus or techniques, new materials, circuits, or other information valuable to investigators in the field will be accepted from time to time. There is no page charge for publication.

Two copies of each manuscript and each figure or table should be sent to:

Dr. Sidney W. Bijou, Child Behavior Laboratory, 403 E. Healey, University of Illinois, Champaign, Illinois 61820

A third copy should be retained by the author. Manuscripts must be double-spaced and may be typed, dittoed, or mimeographed. Each manuscript must include a concise abstract, preferably not exceeding 300 words, which will appear after the title and author's name in the published article. In some papers this abstract will obviate the need for a summary. In general, authors are free to use whatever organization their paper seems to require, but the specifications of the *Publication Manual* of the American Psychological Association are strongly recommended.

Figures and Tables. Figures should be numbered consecutively with Arabic numerals in order of mention in the text; each figure should have a descriptive legend. Legends should be typed together on a separate sheet, double-spaced. All illustrations should be in finished form ready for the engraver. A convenient size for drawings is $8\frac{1}{2} \times 11$ inches (21×27.5 cm). Drawings should be made with India ink on tracing linen, smooth surface white paper, or bristol board. Graph paper if used should be ruled in blue. Grid lines that are to show in the final engraving should be inked in black. Sharp, clear glossy prints of figures are acceptable.

Tables should be typed on separate pages, numbered consecutively with Arabic numerals in order of mention in the text. All tables should have descriptive headings, typed (double-spaced) above the table.

Literature references in the text should be cited by author's surname and the year of publication, e.g., Smith, 1960; Smith *et al.*, 1959; Smith and Jones, 1961, p. 250 (for references to a specific page). If a reference has more than two authors, the citation includes the surnames of all authors at the first mention, but later citations of the same reference include only the surname of the first author and the abbreviations *"et al."* Suffixes *a, b,* etc., should be used following the date to distinguish two or more works by the same author(s) in the same year, e.g., Smith, 1960a, 1960b. Literature citations in the bibliography should be arranged alphabetically according to the surname of the author.

Journal titles should be written out in full according to the form followed in the Psychological Abstracts (Vol. 39).

The desired style of capitalization and punctuation in references is apparent in the following examples:

STAATS, A. W., AND STAATS, CAROLYN K. A comparison of the development of speech and reading behavior with implications for research. *Child Development,* 1962, 33, 831–846.

SPIKER, C. C. Research methods in children's learning. In P. H. Mussen (Ed.), *Handbook for research methods in child development.* New York: Wiley, 1960. Pp. 374–420.

MUNN, N. L. *The evolution and growth of human behavior.* Cambridge: Riverside Press, 1955.

Reprints: Fifty gratis reprints are supplied and additional copies may be ordered.

Negro and White Children

A PSYCHOLOGICAL STUDY IN THE RURAL SOUTH

By **E. Earl Baughman**
W. Grant Dahlstrom
Department of Psychology, University of North Carolina Chapel Hill, North Carolina

1968, 572 pp., $15.00

Reports the scientific conclusions of a unique four year study of racially-segregated students in a rural, economically deprived area of North Carolina.

CONTENTS:

SETTING AND METHODS
Life in Millfield
Procedures Used to Study Millfield's Children
INTELLIGENCE AND ACADEMIC ACHIEVEMENT
Ability Differences between Racial, Sex, and Age Groups
Academic Achievement Differences between Racial, Sex, Age, and Ability Groups
The Relationship of Selected Family Variables to Ability and Academic Achievement.
INTELLECTIVE CHANGE
Changes in IQ Among Young Negro Children
Changes in IQ Among White Adolescents
The Effect of Kindergarten Training upon the Intellective Functioning of Negro and White Children
A Pilot Project To Accelerate the Intellective Development of Four-Year-Old Negro and White Children
PERSONAL ATTRIBUTES
Millfield's Teachers View Their Pupils
Eighth-Grade Children as Assessed by the Minnesota Multiphasic Personality Inventory
Fantasy Behavior: Achievement Motivation
Fantasy Behavior: Other Attributes
Correlates of Peer Popularity
PERSONAL PERSPECTIVES
Mothers of Kindergarten-Age Children Report on Their Children, Training Practices, and Families
Eighth-Grade Children Report on Their Worlds
INTEGRATION AND RECOMMENDATIONS
Some Summary Comments and Observations
Narrowing the Gap

REFERENCES.
APPENDIX
AUTHOR INDEX—SUBJECT INDEX.

AP 2155

Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1

JOURNAL OF Experimental Child Psychology

EDITOR: Sidney W. Bijou

ASSOCIATE EDITORS: Richard S. Bogartz
David Zeaman

EDITORIAL BOARD:

Albert Bandura
Wesley C. Becker
Harry Beilin
Lyle E. Bourne, Jr.
Gordon N. Cantor
Charles C. Clifton, Jr.
Peter Eimas
Robert L. Fantz
Eleanor J. Gibson
Eugene S. Gollin
Willard W. Hartup
Raymond H. Hohle
Herbert Kaye
Tracy S. Kendler
Helen L. Koch
Eugene R. Long
Harry Munsinger
David S. Palermo
Ross Parke
Herbert L. Pick
Hayne W. Reese
Todd Risley
Bryan E. Shepp
Murray Sidman
Einar Siqueland
Patrick Suppes
Thomas M. Trabasso
Paul Weisberg
Joachim Wohlwill
Betty House Zeaman

This Number Completes Volume 8
Volume 8, Number 3, December 1969

ACADEMIC PRESS
New York and London

CONTENTS

JOSEPH W. GALLAGHER. Semantic Relationship and the Learning of Syntactic Word Pairs in Children 411

ROSSLYN GAINES. The Discriminability of Form Among Young Children . 418

RICHARD C. URBANO AND KEITH G. SCOTT. Digit Span, Practice and Dichotic Listening Performance in the Mentally Retarded 432

DENNIS E. MITHAUG. The Development of Cooperation in Alternative Task Situations 443

DAVID A. PARTON AND GREGORY T. FOUTS. Effects of Stimulus-Response Similarity and Dissimilarity on Children's Matching Performance . . 461

LYN W. WICKELGREN. The Ocular Response of Human Newborns to Intermittent Visual Movement 469

GERALD TURKEWITZ, TINA MOREAU, LINDA DAVIS, AND HERBERT G. BIRCH. Factors Affecting Lateral Differentiation in the Human Newborn . . 483

LOUISE S. TIGHE AND THOMAS J. TIGHE. Transfer from Perceptual Pretraining as a Function of Number of Task Dimensions 494

LINDA BILSKY AND LAIRD W. HEAL. Cue Novelty and Training Level in the Discrimination Shift Performance of Retardates 503

EDITORIAL NOTE 512

ANNOUNCEMENT 515

AUTHOR INDEX 516

SUBJECT INDEX 518

Published bimonthly at Exchange Place, Hanover, Pa. 17331 by Academic Press, Inc., 111 Fifth Avenue, New York, N. Y. 10003.

In 1969, Volumes 7 and 8 (consisting of 3 issues each) will be published.

Price: $22.00

Private subscriptions (for subscriber's personal use): $10.00

All correspondence and subscription orders should be addressed to the office of the Publishers at 111 Fifth Ave., New York, N. Y. 10003

Send notices of change of address to the office of the Publishers at least 4 weeks in advance.

Please include both old and new addresses.

Second-class postage paid at Hanover, Pennsylvania 17331

© 1970 by Academic Press, Inc.

In 1970, Volumes 9 and 10 (consisting of 3 issues each) will be published.

Price: $24.00

Private subscriptions (for subscriber's personal use): $10.00

Note to Subscribers to
Journal of Experimental Child Psychology

Volume 9, Number 1, will be sent to all 1969 subscribers. If you have not renewed your subscription for 1970, it is suggested that you do so immediately in order to assure continuity of service.

Academic Press, Inc.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 411–417 (1969)

Semantic Relationship and the Learning of Syntactic Word Pairs in Children[1]

JOSEPH W. GALLAGHER
University of Alabama

The present study examined the influence of semantic consistency (meaningfulness) and anomaly on the learning of three types of syntactic pairs. The results showed that meaningful pairs are learned with fewer errors than anomalous pairs. The difference between meaningful and anomalous pairs increased from kindergarten to grades 1 and 3. This was interpreted as a manifestation of the development of the semantic system. However, there was a decrease between the meaningful and anomalous pairs from grade 3 to grade 5. Meaningful noun-verb and adjective-noun phrases were learned with fewer errors than verb-noun phrases. No differences were found among types of phrases for the anomalous pairs.

Several studies (Marks and Miller, 1964; Miller and Isard, 1963), using adults, have reported that semantically correct sentences are dealt with more readily than anomalous sentences. Miller and McNeill (1969), using children, have presented data that suggest that children 5, 6, 7, and 8 years of age recall more semantically correct sentences than anomalous sentences (no statistical evidence was given). Katz (1966; p. 1961) has noted that anomaly may not only exist on the sentence level, but may exist between any two constituents. This indicates that syntactic word pairs can also be considered as anomalous when the two words are semantically inconsistent. It is expected in the present study that meaningful pairs will be learned faster than anomalous word pairs at all age levels.

Miller and McNeill (1969) also suggested that an interaction existed between the semantic consistency of the sentence (meaningful vs. anomaly) and age level. Miller and McNeill (1969) reported that the difference between the number of meaningful and anomalous sentences correctly recalled increased as a function of age. The present study was designed to examine the influence of semantic relationship on paired-associate (PA) learning of three types of syntactic pairs—adjective-noun (AN), noun-verb (NV), and verb-noun (VN). The present study also extended the age range to include children up to 10½ years of age. It is expected

[1] This research was supported by a University of Alabama Research Committee Grant, Project 531 and by Training Grant HD 124 made to the Psychology Department, University of Alabama.

© 1970 by Academic Press, Inc.

that the difference between the ease of learning the meaningful and anomalous pairs should increase as age increases.

Miller and McNeill (1969) also suggest that only small changes would occur across ages 5–8 for anomalous materials. Furthermore, they suggest that adults (older *Ss*) would recall only slightly more than the 8 year olds they employed. The present study will examine these predictions by using grade-5 *Ss*.

METHODS

Subjects.[2] A total of 144 *Ss* were employed. Thirty-six *Ss* from each of kindergarten (Kg), grades 1, 3, and 5 were used. The mean chronological age (CA) of the children in Kg and grades 1, 3, and 5 was 5 yr. 1 mo., 6 yr. 7 mo., 8 yr. 5 mo., and 10 yr. 6 mo., respectively. The Kg children were randomly selected from a pool of 40 *Ss* attending the University of Alabama kindergarten. The children in grades 1, 3, and 5 were randomly selected from groups of 63, 58, and 65 children, respectively. *Ss* from each age level were randomly assigned to one of four lists (two meaningful and two anomalous). These *Ss* were employed so that one *S* on each list was used before another *S* was used from the same list. This was done for all age levels. All *Ss* at one age level were used before another age level was started.

Lists.[3] Four lists of nine word pairs were constructed. Lists I and II consisted of meaningful syntactically correct word pairs such as "chairs fell" and "loud boy." Lists III and IV consisted of anomalous word pairs such as "chair talks" and "loud air." Each list consisted of three AN, three NV, and three VN phrases. All stimuli were taken from the Palermo-Jenkins (1963, 1964) norms for grade 5. The free association strength (FAS) values between all stimuli and responses was 0.0% (Palermo and Jenkins, 1963, 1964). Interitem associations among the stimulus items was 0.0% (Palermo and Jenkins, 1963, 1964). Interitem associations among the response items were judged to be minimal by the *E*. All words were at least 30 on the Thorndike-Lorge (T-L) J scale (Thorndike and Lorge, 1944). Word frequency was approximately the same in all four lists for both the stimulus and the response words. All words employed were correctly defined by independent groups of 10 children from Kg and 10 from grade 1.

Adult judges agreed 100% as to which pairs were meaningful and nonmeaningful. It was assumed that adult judges would have a complete

[2] The author thanks Dr. Jacqueline Davis, Director, University of Alabama Nursery School and Roger Ballard, Principal, Vestavia Elementary School, Tuscaloosa County Schools.

[3] The lists of word pairs are available upon request to the author.

set of semantic markers for the words employed. Thus, an increase in learning rate found between the meaningful and anomalous pairs across age level could be attributed to the growth of the child's semantic system.

Apparatus and procedure. A Wolensak tape recorder was used to present the materials. The presentation rate employed a 4-second anticipation interval during which the stimulus word was said at the beginning of the interval, followed by a 4-second interval at the start of which *S* heard both the stimulus and response repeated together. An 8-second intertrial interval was used during which the *S* heard nothing.

The *E* informed the *S*s that they would hear words on the tape recorder and their job was to learn which words went together. Each *S* was given practice with two pairs—one meaningful and one anomalous pair. During Trial 1, which was a familiarization trial, *S*s were instructed to say aloud everything they heard. After Trial 1, *S*s were instructed to say aloud only the response word when they heard the stimulus word. Six random orders of presentation were used for each list to prevent serial learning. *S*s were run for 15 trials after the familiarization trial.

RESULTS

Table 1 presents the means and standard deviations for the three types of phrases and two types of pairs at four age levels. It can be seen that the difference between the meaningful and anomalous pairs increase as age

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR THREE TYPES OF PHRASES AND TWO TYPES OF PAIRS AT FOUR AGE LEVELS[a]

	Meaningful							
Grade level	NV		AN		VN		Total	
	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
Kg	8.83	3.62	9.00	3.74	14.83	4.68	32.66	8.57
1st	2.67	1.97	2.75	2.55	4.58	2.44	10.00	5.66
3rd	3.42	3.84	2.75	1.96	5.08	2.47	11.25	4.28
5th	0.83	0.90	1.00	1.20	2.67	1.37	4.50	2.63
Total	3.94	5.05	3.88	5.25	6.79	7.84		
	Nonmeaningful							
Kg	11.33	9.43	12.67	4.46	11.83	3.15	35.83	12.89
1st	8.50	2.10	10.14	5.71	10.58	6.12	29.22	10.41
3rd	8.00	3.24	8.76	4.14	8.17	4.17	24.93	9.79
5th	3.92	2.65	5.08	3.48	3.33	2.46	12.33	6.49
Total	7.94	7.86	9.16	7.11	8.48	6.85		

[a] The means and standard deviations are for the nontransformed data.

increases, except at grade 5 where there is a decrease. It can also be seen that, for the meaningful pairs, there is no difference between the NV and AN phrases while these two types of phrases are learned with fewer errors than the VN phrases. For the anomalous pairs, fewer errors were made on the NV and AN phrases, but a greater number of errors were made on the AN than on the VN phrases. In general, performance increases as age increases. However, for the meaningful pairs, grade 1 made slightly fewer errors than grade 3.

Due to apparent increase in variance with an increase in age, a F_{max} test was conducted. The F_{max} was 77.61 ($p < .01$) indicating a lack of homogenity of variance. In view of this, the transformation $x^1 = \sqrt{x} + \sqrt{x+1}$ (Winer, 1962, p. 220) was applied to the data. The F_{max} on the transformed data was 2.31 ($p > .05$). All of the statistical tests conducted were done on the transformed data.

An analysis of variance was conducted on the number of errors across Trials 2–16. The analysis consisted of two-between-*S* factors, Grade Level (Kg, grades 1, 3, and 5), and Type of Pairs (meaningful and anomalous); and one-within-*S* factor, Type of Phrase (NV, AN, and VN).

The main effect of Type of Pair was significant ($F(1/136) = 116.676$, $p < .01$) indicating that the meaningful pairs were learned with fewer errors than the anomalous pairs. The main effect of Type of Phrase was also significant ($F(2/272) = 5.515$, $p < .01$). Individual comparisons were conducted using the Tukey (b) test. No significant ($p > .05$) difference was found between the NV and AN phrases. Both the NV and AN phrases were learned with significantly ($p < .05$) fewer errors than the VN phrases. The main effect of Grade Level was significant ($F(3/136) = 18.402$, $p < .01$).

The Grade Level $\times$ Type of Pair interaction was significant ($F(3/136) = 3.853$, $p < .01$) indicating that the influence of the Type of Pair was not the same at each Grade Level. The difference between the meaningful and anomalous pairs increases from Kg to grades 1 and 3, but this difference decreases as one goes from grade 3 to grade 5. Individual comparisons within Grade Level were conducted using the Tukey (b) test. The results show that the difference between the meaningful and anomalous pairs was not significant ($p > .05$) at the Kg level. For grades 1, 3, and 5, the meaningful pairs were learned with significantly ($p < .05$) fewer errors than the anomalous pairs. Individual comparisons using the Tukey (b) test were also conducted across Grade Level for each of the two types of pairs. For the meaningful and anomalous pairs, grades 1, 3, and 5 made significantly ($p < .05$) fewer errors than the Kg children. For both types of pairs, grade 5 made significantly ($p < .05$) fewer errors than grades 1 and 3, while first and third grades did not differ significantly ($p < .05$).

The Type of Phrase × Type of Pair interaction was significant (F (2/272) = 4.829, $p < .01$). This interaction appears to be due to the fact that for the meaningful pairs the NV and AN pairs are learned with fewer errors than the VN pairs, while this same relationship is not found for these three Types of Phrases for the anomalous pairs. Individual comparisons were conducted with the Tukey (b) test. For the meaningful pairs, the NV and AN were learned with significantly ($p < .05$) fewer errors than the VN, while no significant ($p > .05$) differences were found between the NV and AN pairs. For the anomalous pairs, no significant differences were found among the three Types of Phrases. None of the other interactions were significant.

DISCUSSION

The overall finding that semantically consistent word pairs are learned with fewer errors than anomalous pairs supports and extends the results of other investigators who have examined the influence of semantic consistency and anomaly with adults (Miller and Marks, 1964; Miller and Isard, 1963) and children (Miller and McNeill, 1969) to cover other anomalous units.

The significant Grade Level × Type of Pairs interaction was expected and to some extent supports the results of Miller and McNeill (1969). In the present study, it was found that from Kg to grades 1 and 3, there was an increase in the difference between the number of errors made on the meaningful and anomalous word pairs. Thus, indicating that as age increases, children are better able to take advantage of semantic consistency.

Miller and McNeill (1969) suggested there would be little change across ages 5–8 for anomalous sentences and that adults would recall "slightly more" than the 8 year olds for both the semantically consistent strings and the anomalous strings. The results of the present study indicated that grade 5 ($\bar{X}$CA = 10 yr. 6 mo.) children make fewer errors than those in grade 3 ($\bar{X}$CA = 8 yr. 5 mo.) for both the meaningful (semantically consistent) and anomalous word pairs employed. The difference between the mean number of errors for grades 3 and 5 for the meaningful pairs was 6.75 and for the anomalous pairs it was 12.60. Thus, in PA learning, examining the influence of the semantic relationship between the words, children beyond 8 years of age appear to show increasing abilities to take advantage of both semantic and syntactic properties of word pairs. The present study employed a typical PA anticipation task while Miller and McNeill presented children with sentences masked by noise. The children then had to recall the sentences. Since the children in the Miller and McNeill (1969) did not correctly recall a high percentage of anomalous sentences, it is suggested that the task may have

been too difficult for children (5–8 years) to show increases in learning when anomalous sentences are employed in a recall task.

The finding that AN phrases are learned with fewer errors than VN phrases supports the results of Gallagher (1968) who employed AN and VN phrases as response units in a PA task with adult *Ss*. The finding that the NV phrases were learned with fewer errors than that of the VN phrases has not been reported in past research. Presently, there is no apparent explanation for this differential responding to the different types of syntactic pairs. Some specific characteristics of the stimuli that may have some influence on the ability of a child to associate words were examined. Palermo (1965) reported, in learning nonassociated paradigmatic pairs, that word pairs that consisted of stimuli with steep associative hierarchies (i.e., the FAS value of the primary response is high while the FAS of the next four highest responses is much lower) are learned faster than nonassociated pairs with stimuli with flat associative hierarchies (i.e., when the FAS value of the primary response and the FAS of the next four highest is similar). In the present study, the mean stimulus associative hierarchy was calculated for the stimuli employed in the AN, NV, and VN phrases. No differences were found between the types of phrases for the mean FAS of the primary responses for children in grade 5.

In examining the stimulus associative hierarchies of stimulus words used in the present study, it was found that the mean number of different associates given to verbs (Mean = 62.5) was greater than that given to nouns (Mean = 47.3) and adjectives (Mean = 43.8), while a small difference was found between the number of associates elicited by the nouns and adjectives. To some extent, the results with the NV and AN agree with Glanzer (1962) who found that, across trials for the same *S*, nouns and adjectives elicit fewer responses than verbs. Furthermore, when nouns and adjectives were used as stimuli with nonsense syllables as responses in PA learning, the mean number of errors for pairs with nouns and adjectives was less than that for pairs with verbs. However, Glanzer (1962) also reported that the mean number of errors for pairs with adjectives was less than that of verbs. This is not in complete agreement with the results of the present study. Thus, it appears that the differential learning of syntactic units may not be attributable to specific stimulus characteristics.

The Type of Pair × Type of Phrase interaction indicated that differences between types of phrases was found with the meaningful pairs but not with the anomalous pairs. Since the stimuli were the same for both types of pairs, this suggests that differential responding to various phrases should be examined in view of the semantic relationship between the

words. Again, this suggests that specific stimulus characteristics may have little differential effect on the learning of syntactic pairs since the same stimuli produce different results.

REFERENCES

GALLAGHER, J. W. An investigation of the effect of free association strength on the learning of syntagmatic word pairs. *Journal Verbal Learning Verbal Behavior,* 1968, **7**, 726–729.

GLANZER, M. Grammatical Category: A rote learning and word association analysis. *Journal Verbal Learning Verbal Behavior,* 1962, **1**, 31–41.

KATZ, J. J. *The Philosophy of Language,* New York: Harper, 1966.

MARKS, L. E., AND MILLER, G. A. The role of semantic and syntactic constraints in the memorization of English sentences. *Journal Verbal Learning and Verbal Behavior,* 1964, **3**, 1–5.

MILLER, G. A., AND MCNEILL, D. Psycholinguistics. In G. Lindzey and E. Aronson (Eds.), *Handbook of Social Psychology,* Vol. III: The individual in a social context. Reading, Mass.: Addison-Wesley, 1969.

MILLER, G. A., AND ISARD, S. Some perceptual consequences of linguistic rules. *Journal Verbal Learning Verbal Behavior,* 1963, **2**, 217–228.

PALERMO, D. S. Effects of stimulus free association hierarchy on paired associate learning of unassociated words. Paper presented at meeting of Midwestern Psychological Association, Chicago, May, 1966.

PALERMO, D. S., AND JENKINS, J. J. *Word association norms: Grade school through college.* Minneapolis: University of Minnesota Press, 1964.

PALERMO, D. S., AND JENKINS, J. J. *Free association responses to primary responses of the Palermo-Jenkins word association norms for grade school children.* Res. Bull. #37, Department of Psychology, The Pennsylvania State University, 1963.

THORNDIKE, E. L., AND LORGE, I. *The teacher's word book of 30,000 words,* New York: Teachers College, Columbia University, 1944.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY 8, 418–431 (1969)

The Discriminability of Form Among Young Children[1]

ROSSLYN GAINES
University of California, Berkeley

The ability of young children to discriminate accurately forms which varied in complexity, line type, and structure was investigated. Variations in complexity or number of sides, in form perimeter from linear to curvilinear to combined linear-curvilinear, and in structure from symmetrical to asymmetrical (compact-dispersed) were increasingly difficult to discriminate as measured by both errors and latency of response. Sex and IQ were unrelated to performance. It was found that nursery, kindergarten, and first grade children were all above chance in performance, older children being superior. The relations of these data to developmental perceptual theory are discussed.

When children are asked to select stimuli which vary in physical dimensions, two types of information emerge: first, a measure of a child's skill in discerning the variations; second, relations between stimuli characteristics and skill.

Regarding the first set of data, a current view (Bruner, 1966) is that a child's perception is "stuck, autistic, diffuse, concrete, and unsteady." In contrast, approximately 40 years ago, a child's perception was viewed as discrimination "to the limits of sensory acuity, seizing each thing in its uniqueness, noting every hair and flea of the particular dog" (Brown, 1958). Some current studies of infant (e.g., Bower, 1966), child (e.g., Pollack, 1963) and animal (e.g., Ganz and Wilson, 1967) perception suggest that this earlier view of children's "sensory acuity" may have some validity. Since these studies did not give information on form discrimination, the present study was undertaken.

[1] This research was supported by USPHS National Institute of Mental Health grant MH11234 to the writer. The writer expresses gratitude for the continuing cooperation of Francis Boarders and Constance Ackerman of Beauvoir National Cathedral School and Peter Rice of Sidwell Friends School, Washington, D. C. Mrs. Epp Miller, Betty Bond, and Lynn Skelfo, George Washington University, were responsible for careful testing and tabulation of the data, and William Mallory, University of California, Berkeley, assisted in the analysis of the data. Dr. Tom Trabasso, Princeton University, gave a useful critique of the manuscript. Dr. Norman Anderson, University of California, San Diego, has been helpful in suggesting methods of statistical analysis. Portions of this paper were presented at the American Psychological Association meeting, Washington, D. C., 1967.

The metrics of children's responses to form dimensions have not been delineated, although those for adults are quite extensive (Michels and Zusne, 1965; Brown and Owen, 1967). For adults, differences such as jaggedness, compactness, and elongation clearly alter speed of responding. The effect of these variables on responses of young children has not been intensively studied despite research (e.g., Suchman and Trabasso, 1966) showing stimulus variables do serve young children as cues in learning.

Studies of children's responding to form stimuli frequently use looking time as a dependent variable. In general, developmental differences in response to varying form structures are examined. For example, Munsinger and Kessen (1966) found that older children estimated asymmetric figures better than symmetric figures, but young children estimated symmetric and asymmetric figures equally well. However, Crudden (1941) found that young children had more difficulty differentiating asymmetrical embedded figures than symmetrical embedded figures. This dilemma was in part made more explicable by Forsman's (1967) finding that younger *Ss*' response time to symmetric and asymmetric figures varied with the task: on an estimation of sides task, no differences occurred, but on a matching task, response latencies to asymmetric figures were longer. Again, in the Munsinger *et al.* (1964, 1966) studies, children estimated and categorized 10-sided figures better than 5- and 20-sided figures. However, Thomas (1966) repeated the latter study with different results. Among 6- to 12-year-old children, attention to complexity increases as the number of sides increases, with 40-sided figures looked at longest.

One explanation of the discrepant findings might reside in Attneave and Arnoult's (1956) phrase, "Shape is a multidimensional variable." More reliable indices of children's responses to particular dimensions could be obtained if choice responses are measured on several form dimensions simultaneously. The present study measured children's latencies and error rates in response to several dimensions of form.

First, latency was chosen as a measure because adult discrimination latency (Brown and Andrews, 1968) can be used as a distance measure in multidimensional scaling. Further, the subjective dimensions can be psychophysically predicted from form metrics. The present study investigates whether the finding that latency increases as complexity increases is also characteristic of children's responses. On the basis of adult data (Brown and Andrews, 1968), young children should also take longer to differentiate figures which have more sides, are structurally complex, have a complex line type, and have a smaller magnitude of change. In addition, variation in contour produced by either subtracting from or adding to the total area of the base figure is examined (Fig. 2). The general thesis

is that children's form perception resembles that of adults in that similar physical stimulus characteristics are seen as complex and that as forms become more complex and the change is still discriminable as an increase in complexity, latency increases.

Second, this study varied discriminability of differences between forms, such as small size variations (1–3 mm), since looking time could be related to the child's capacity to discriminate form variations, as well as to either the child's choice or the stimulus complexity. Children are often asked to find small form or size variations among forms but without any systematic attempt by the experimenter to scale the difficulty level of the stimulus variables used in the discrimination task. This leads to error source being attributed to sample or experimental conditions without consideration of the effect of stimulus variables on performance. Finally, the oddity problem method was selected for this study, since it permits measures of small form variations. The general thesis is that as sides, structural complexity, and complexity of line type are still discriminable as increases in complexity, the error rate increases and reflects the complexity of the figure as accurately as latency measures.

METHOD

Stimulus Materials

The first set of stimulus materials consisted of 100 form-oddity problems. The level of difficulty was based on reports of young children's perceptual skills (e.g., Hill, 1965; Osler and Kofsky, 1965), and almost half of the problems were designed to be more difficult than young children could be expected to solve.

Pilot testing was done on the youngest age group (4 years) and successive sets of increasingly difficult problems were pretested. The first *Ss* tested made few errors, and their high scores were attributed to either known perceptual skill or superior intelligence. After several *Ss* failed to make any appreciable number of errors, the problems were discarded and a new series of more difficult problems was designed.

The second set of stimulus variables and their dimensions are listed below as shown on Figure 1. Within dimensions, the easier values are listed first: (a) two levels of Sidedness: 4 and 8 sides; (b) three levels of Structure: symmetrical compact (e.g., a square); asymmetrical-compact (e.g., a trapezoid); and asymmetrical dispersed (e.g., an arrow-type shape) (structure includes the variables of symmetry, compactness, and jaggedness); (c) three Line Types: linear, curvilinear, and combined linear-curvilinear; (d) three Size Changes were selected: 6-, 3-, and 1-mm changes along the perimeter of figures approximately 1 square inch. The

6-mm change was made by using two 3-mm changes on each odd figure. The area range for all figures in the series, as measured by a planometer, is 1.4 .80 inches. The size changes varied shape several ways and varied structure two ways. The shape and structural variations comprised: (1) side length extended, (2) side length shortened, (3) corner cut off, (4) curvature increased, and (5) curvature decreased. A change in one side changes an adjacent side and the above categories describe the origin of the change, rather than the total physical change in shape.

All figures were drawn by a professional draftsman to specifications. The variations in size were randomized among figures and addition to or subtraction from the area of the figure was counterbalanced within each of the dimensions (a–d above). The basic figures used are shown in Figure 1. An example of the oddity problem format is shown in Figure 2. In general, four same figures were compared with one figure in each of the three size changes, with an addition and a subtraction in each of the 3- and 1-mm changes. Position of changes for the four corner figures were counterbalanced within basic figure groups.

Prior to the final statistical analysis it was found that there was no significant difference in error rate between 6- and 3-mm figures; the probability of error was .09 for both problem sets (Suchman, 1967). Since there was no difference in error rate and since the 6-mm problems did not include the direction of change variable, these problems were not included in the final analysis. Thus, each stimulus item represents one level of sidedness, one structural level, one line type, one size change, either added or subtracted, in a $2 \times 3 \times 3 \times 2 \times 2$ design or 72 test items.

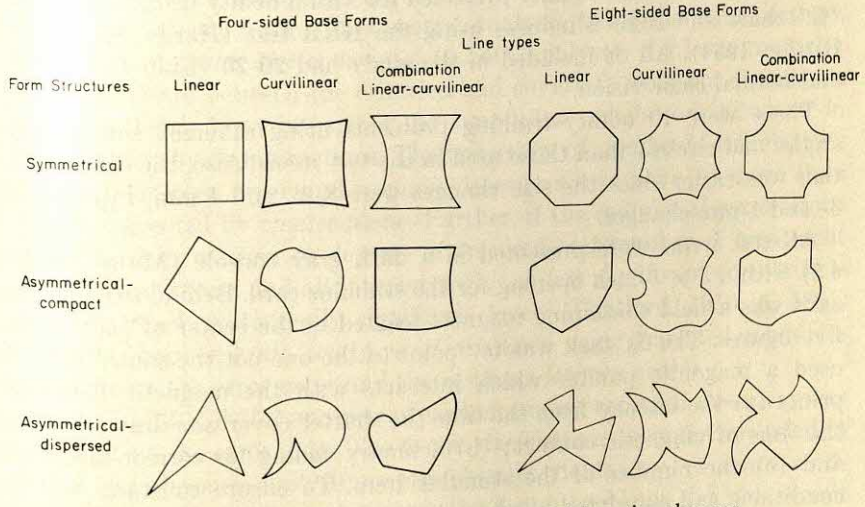


FIG. 1. Base form for each of the oddity problem size changes.

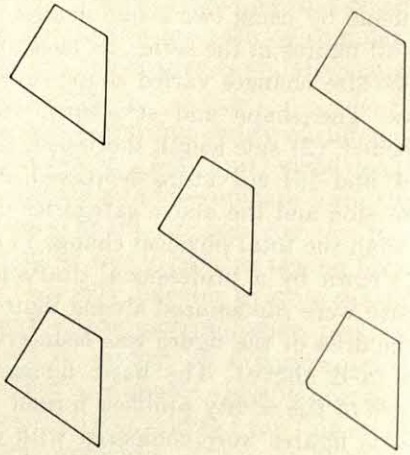


FIG. 2. A form oddity example. In this example, the odd item is in the lower right-hand corner. The form variables are four-sidedness, compact-asymmetrical structure, and linear line type. The size change is made by removing one millimeter from the length of the base line. Original display size is 5 × 5 inches.

Procedure

The sample consisted of thirty *S*s, ages 4.6–7.6 years, from a private school in Washington, D. C. There were 15 girls and 15 boys from nursery school, kindergarten, and first grade. The mean IQ was 135. The IQ range was 109–159. The majority of *S*s were children of upper class parents as measured by income, occupation, and education.

Children were individually pretested for visual acuity using a tumbling "E" chart and color blindness using the HRR test (Hardy, Rand, and Rittler, 1954). All *S*s included in the study had 20-20 vision (corrected) and normal color vision.

There were 10 oddity-training problems using different shapes (triangles and circles) than those used in the test items. Also, the discrimination was easier since the size changes were 8, 6, and 4 mm, rather than 3- and 1-mm changes.

All test items were presented in a dark gray console (Munsell value 4.5) with a 5 × 5-inch opening for the stimulus card. Behind the stimulus card was a field containing magnets located in the center of each of the five figures. The *S*s' task was to "point to the one not the same." The *S*s used a magnetic pointer which interacts with the magnetic field and prints (a) the latency from the time the shutter cover was drawn back to the time of magnetic contact; (b) a binary coding for correct-incorrect; and (c) the number of the stimulus item. To ensure constant lighting conditions and simulated northern light, the stimulus cards were lit with

a MacBeth daylight lamp. The incident light at the screen was 90 foot-candles, as measured with a Wesson light meter. Each *S*'s seating height was adjusted so that the center of the console display was at his eye level.

Since the important variable was task performance, the *S*s were given, systematically, four types of reinforcement. First, *S* preselected a toy from an array of several five-cent toys (e.g., magnets, gliders, necklaces, yo-yos). The *S* was told that he could have the toy when he finished playing the game. Second, *S* automatically received a marble for each correct response. Third, *S* was given corrective feedback on each incorrect response; that is, the part of the form which was different was pointed to by *E*. Fourth, *S* was verbally praised during training and at the end of the session.

The *S*s had two sessions 1 week apart on the 72 form oddities. In the first session, the *S* had 6 training items and 24 oddity problems randomly arranged except for the exclusion of 1-mm size changes. In the second session, the *S* had 4 training items and 48 problems randomly arranged except for the exclusion of 6-mm size changes.

RESULTS

Two analyses were made of the data. First, the association between error and response latency scores and the form variables were investigated. Second, the effects of sex, IQ, and grade level on errors were analyzed.

First, the error rate was assessed by a $2 \times 3 \times 3 \times 3 \times 2$ analysis of covariance using repeated measures. In one analysis, the response latency scores were the covariate and adjusted the errors for the linear influence of response latency scores. In a second analysis, using the same method, the error terms were the covariate. In both analyses, the large majority of the correlates between the criterion and covariate were not significant at the .05 level. Most important, no significant correlations occurred in the treatment by *S*s interactions. The frequency of significant correlations (4 of 62) between error and response latency scores were no more than could be expected by chance alone. Further, if the individual correlations are assessed in relation to degrees of freedom and number of treatment effects (main effects and all interactions), the few significant correlations no longer reach statistically significant levels (Guilford, 1956, p. 538). Therefore, two separate analyses of variance using repeated measures were done on the error rates and the response latency scores.

A $2 \times 3 \times 3 \times 3 \times 2$ analysis of variance using repeated measures was performed on the error data. The 6-mm size changes were not included in this analysis because first, 3-mm size changes yielded as few errors as 6-mm problems ($t = .14$), and second, the 6-mm problems did not permit

comparison of direction of change (each 6-mm change consisted of one 3-mm change adding to, and one 3-mm change subtracting from, the total area).

The analysis was made using the UCLA Bi-Med program 08V (Bennett and Franklin, 1954) and then discarding the pooled error term. An error term for each main effect was derived by the interaction of the effect with subjects, producing F ratios adjusted for repeated measures. The analysis of variance on errors yielded significant F ratios for all main effects except direction of size changes. Table 1 shows the mean error rates and mean latencies for main effects of the analysis. Individual analyses of the tri-levels within the main effects were performed by t tests for correlated means. As seen on Table 1, the four-sided discriminations were easier than the eight-sided discriminations. Again, symmetrical-compact figures had lower error rates than asymmetrical-compact figures ($t = 3.67$, $p < .01$) and the asymmetrical-dispersed figures produced the highest error rate when compared with symmetrical-compact and asymmetrical-compact figures ($t = 3.28$, $p < .01$; $t = 5.76$, $p < .001$, respectively). Line type was not effective alone but the analysis clearly shows that linear-curvilinear combined figures were easier to discriminate than linear or curvilinear figures ($t = 2.43$, $p < .05$; $t = 5.35$, $p < .001$, respectively). There was no significant difference in error rate between linear and curvi-

TABLE 1

MEAN ERROR RATE AND MEAN LATENCY SCORES OF 30 Ss ON FORM DIMENSIONS OF ODDITY PROBLEMS

Dimensions	No. of problems	Mean error	Mean latency (seconds)
Structure			
Symmetrical compact	24	2.59	8.32
Asymmetrical compact	24	4.30	9.66
Asymmetrical dispersed	24	5.42	13.92
Line type			
Linear	24	4.27	11.02
Curvilinear	24	5.06	13.62
Combined Linear-Curvilinear	24	3.00	7.25
Sides			
Four sides	36	4.86	8.30
Eight sides	36	7.45	12.96
Size of oddity change			
Three mm	36	3.53	6.59
One mm	36	8.78	14.67
Direction of change			
Area added	36	6.50	11.05
Area subtracted	36	5.83	10.21

linear figures. The 3-mm size change had a lower error rate than 1-mm size changes. There was no significant difference in direction of change. It is possible that larger size variations than 1 and 3 mm would produce a more observable effect on direction of change.

The significant F ratios are shown in Table 2. In interactions as complex as the present analyses, some trends can be indicated. These trends cannot be taken to "explain" results without further research. Only one variable, line type, was significant in interaction with all other main

TABLE 2
ANALYSIS OF VARIANCE OF FORM ODDITY ERRORS

Source[a]	*df*	*MS*	*F* ratios[a]
A (Sidedness)	1	2.82	16.59
Subjects	29	.17	
B (Structure)	2	2.54	21.17
Subjects	58	.12	
C (Line Type)	2	1.36	9.71
Subjects	58	.14	
D (Size Change)	1	11.56	96.33
Subjects	29	.12	
E (Direction of Change)	1	.19	
Subjects	29	.18	
A × C	2	1.37	11.42
Subjects	58	.12	
B × C	4	.48	4.36
Subjects	116	.11	
C × D	2	1.02	9.27
Subjects	58	.11	
C × E	2	.81	6.23
Subjects	58	.13	
A × C × E	2	1.27	8.47
Subjects	58	.15	
B × C × E	4	2.52	25.20
Subjects	116	.10	
B × D × E	2	.59	7.38
Subjects	58	.08	
C × D × E	2	.50	5.55
Subjects	58	.09	
A × B × C × E	4	.28	3.50
Subjects	116	.08	
A × B × D × E	2	.34	3.09
Subjects	58	.11	
B × C × D × E	4	.73	6.64
Subjects	116	.10	

[a] The F ratios for the main effects and for all interactions significant at $<.01$ are reported.

effects. No other significant 2-way interactions occur. Examining the error rate means corresponding to each of the main effects in the three line types, it appears that when the line type was linear-curvilinear combined, the error rate was lower within sidedness, within structure, within size of change, and within additive or subtractive changes.

The significant three-way and four-way interactions were less clear: sidedness was statistically significant in only one three-way interaction, while line type and additive vs. subtractive were significant in three three-way interactions. All variables were represented in the significant five-way interactions and require a more complex explanation than afforded by this data. In general, examining error rate means corresponding to each of the five main effects on Table 2, it appears that the easiest (lowest error rate) combination of stimulus characteristics was four-sided, symmetrical-compact, linear-curvilinear combined line type discriminations where a 3-mm size change was subtracted from the standard. The highest error rate was in response to eight-sided, asymmetrical-dispersed, curvilinear discriminations where a 1-mm size change was added to the standard.

The analysis of variance performed on the response latency scores had the same structure as the analysis of variance performed on error rate: namely, a $2 \times 3 \times 3 \times 3 \times 2$ analysis using repeated measures. Again, the analysis of variance yielded significant F ratios for all main effects except direction of size change. Table 3 shows the significant F ratios in the analysis of variance and Table 1 the mean latency scores for each of the variables within the main effects. Not only are the same effects significant with respect to time to solution as are significant with respect to error rate, but the direction of the intra-variable differences are identical; that is, response time to four-sided figures was less than to eight-sided figures, response time to symmetrical-compact was less than to asymmetrical-compact figures ($t = 2.10$, $p < .05$). Responses to asymmetrical-dispersed have a longer latency than to symmetrical or asymmetrical-compact figures ($t = 8.19$, $p < .001$; $t = 5.24$, $p < .001$, respectively). In this analysis, responses to linear figures took less time than to curvilinear figures ($t = 3.26$, $p < .01$), but responses to linear-curvilinear combined line type took less time than to linear and curvilinear line type figures ($t = 5.86$, $p < .001$; $t = 7.89$, $p < .001$, respectively). There were lower response latencies to 3-mm size changes than to 1-mm size changes. There were no significant differences in direction of change.

The significant two-, three-, and four-way interactions require further investigation. One major support for continuing investigation is that identical response patterns are observed in error rate and latency measures. Latency analysis is consonant with error analysis as follows: least time was required to respond to a four-sided, compact-symmetrical, linear-curvilinear combined line type figure where a 3-mm size change

TABLE 3
ANALYSIS OF VARIANCE OF FORM ODDITY RESPONSE LATENCY SCORES

Source[a]	*df*	*MS*	*F* ratios[a]
A (Sidedness)	1	11,727.28	46.04
Subjects	29	254.67	
B (Structure)	2	6,168.34	33.39
Subjects	58	184.74	
C (Line Type)	2	7,379.67	36.09
Subjects	58	204.46	
D (Size Change)	1	35,194.69	70.14
Subjects	29	501.71	
E (Direction of Change)	1	374.17	
Subjects	29	96.54	
A × C	2	3,655.55	20.81
Subjects	58	175.67	
A × D	1	2,074.96	13.19
Subjects	29	157.31	
A × E	1	2,275.53	17.74
Subjects	29	128.27	
B × C	4	1,519.43	12.81
Subjects	116	118.62	
B × E	2	859.08	6.31
Subjects	58	136.05	
C × D	2	2,586.54	14.38
Subjects	58	179.93	
C × E	2	3,667.99	20.14
Subjects	58	182.15	
A × B × C	4	412.37	3.60
Subjects	116	114.63	
A × C × E	2	2,144.51	13.57
Subjects	58	158.07	
A × D × E	1	872.67	9.70
Subjects	29	89.97	
B × C × E	4	6,905.01	33.66
Subjects	116	205.11	
B × D × E	2	2,441.31	17.11
Subjects	58	142.72	
C × D × E	2	2,035.62	16.67
Subjects	58	122.19	
A × B × C × D	4	343.64	3.55
Subjects	116	96.85	
A × B × C × E	4	983.44	6.89
Subjects	116	142.80	
A × C × D × E	2	1,301.82	9.41
Subjects	58	138.30	
B × C × D × E	4	3,522.16	26.35
Subjects	116	133.66	

[a] The *F* ratios for the main effects and for all interactions significant at <.01 are reported.

was subtracted from the standard. Most time was required to respond to an eight-sided, asymmetric-dispersed, curvilinear line type figure where a 1-mm size change was added to the standard.

The relationship between grade level and error rate was compared by t test. As expected, the higher the grade, the lower the error rate. Thus, first grade children make fewer errors than kindergarten children ($t = 2.90$, $p < .01$) and kindergarten children make fewer errors than nursery school children ($t = 3.86$, $p < .01$). This finding is no way contradicts the analysis of variance showing that young children, ages 4.6–7.6 years, can make exceedingly small perceptual discriminations. The form oddity error rate among all *Ss* indicates that the probability of a *S* making an error on level 1 (6-mm size change) was .09, on level 2 (3-mm size change) was .09, and on level 3 (1-mm size change) was .23; all above chance level. The form oddity percentage of errors per grade level was 7% for first graders, 13% for kindergarten, and 30% for nursery schoolers. The percentage correct is well above chance level.

In addition, the Pearson r between *Ss'* IQ and error rate was —.02, indicating no relation between error rate and IQ (109–159 range). Last, comparison of the mean error rates of boys and girls were not significant ($t = 0.89$).

DISCUSSION

In this study, young children were found to be surprisingly skilled in discriminating small form details. Replication of this study with a larger and more representative sample is currently under way. The stimuli will include 3-, 2-, and 1-mm, rather than 6-, 3-, and 1-mm changes in the odd figure.

Though the present results are novel, they are in accord with implications of some research suggesting that younger children differentiate between perceptual stimuli with considerable skill. For example, Gibson and Yonas (1966) studying visual search behavior, discovered that second-grade children were as fast as older children and adults in scanning one and two targets, despite predictions to the contrary. Pollack (1963) found that, among children 8–12 years old, a tachistoscopic presentation of a line bisecting a field was perceived at lower thresholds of illumination by the younger children than by the older children. Last, Wescott and Tolchin (1966) report that in a study of identification of incomplete pictures, "Some nursery school subjects do better than some college subjects and almost half of the first grade scores are in the range of the college scores." The authors conclude that "the lack of discontinuity in skill is related to personality variables." It is just as feasible to

relate high scores in visual task performance to the ability to differentiate small variations within the form stimulus dimensions.

Another source of relevant data is the study of children's attention. To the writer's knowledge, only one study (Ricciuti, 1963) has dealt with children's attention to geometric detail vs. geometric form. Ricciuti found that preferred cue utilization was stable across several tasks and that for children ages 3–7 years, approximately 20% of all judgments were based on stimulus detail rather than geometric form. It is certainly clear that for the stimulus materials used, form was a more compelling source of attention than detail. But there were "some children at every age level for whom the predominant cues are the stimulus details rather than over-all shape." Moreover, young children (Beilin, personal communication), on a sameness identity task, continually found small detail differences which were irrelevant to the concept being studied but apparent to the child. Podell (1965) also refers to such attention to "irrevelant" cues, making it necessary, on the basis of a pilot study, to redesign test stimuli.

Children's perceptual skills currently are described as interfering with their ability to respond veridically (Beilin, 1964), syncretic (Werner, 1957), and nonfunctional in cognitive processes (Piaget, 1962). It is also apparent that there are areas of perceptual development where young children have difficulty, for example: the perception of diagonals (Rudel and Teuber, 1963), the discrimination between right and left orientations (Sekuler and Rosenblith, 1964), and the confusions between b and d, p and q (Davidson, 1935). However, despite young children's multiple errors and multiple difficulties in perceptual discrimination, there is also no question that they seem to have excellent, if not superior, skill in discerning small detail variations. Earlier writers (e.g., Stern, 1924; Welch, 1939) described this fine perceptual skill of young children, but generally dismissed this skill as nonrelevant to, or interfering with, cognitive development. However, despite multiple theories, there are no data demonstrating that perceptual skill inhibits the development of cognitive skill. In any event, inadequate information regarding the function of this skill does not counterindicate recognition of young children's perceptual skill.

REFERENCES

ATTNEAVE, F., AND ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychological Bulletin*, 1956, **6**, 452–471.

BEILIN, H. Perceptual-cognitive conflict in the development of an invariant area concept. *Journal of Experimental Child Psychology*, 1964, **1**, 208–226.

BENNET, C. A., AND FRANKLIN, N. L. *Statistical analysis in chemistry and the chemical industry*. New York: Wiley, 1954.

Bower, T. G. R. Visual perception of human infants. Institute of Human Learning colloquium, University of California, Berkeley, 1967.

Brown, D. R., and Andrews, M. H. Visual form discrimination-multidimensional analyses. *Perception and Psychophysics,* 1968, 3, 401–406.

Brown, D. R., and Owen, O. H. The metrics of visual form: Methodological dyspepsia. *Psychological Bulletin,* 1967, **68,** 243–259.

Brown, R. How shall a thing be called? *Psychological Review,* 1958, **65,** 14–21.

Bruner, J. On cognitive growth: I. In J. Bruner and R. Olver (Eds.), *Studies in cognitive growth.* New York: Wiley, 1966.

Crudden, C. H. Form abstraction by children. *Journal of Genetic Psychology,* 1941, **58,** 113–119.

Davidson, H. P. A study of the confusing letters, "b, d, p, q." *Journal of Genetic Psychology,* 1935, **47,** 452–465.

Forsman, R. Age differences in the effects of stimulus complexity and symmetrical form on choice reaction and visual search performance. *Journal of Experimental Child Psychology,* 1967, **5,** 406–429.

Ganz, L., and Wilson, P. D. Innate generalization of a form discrimination without contouring eye movements. *Journal of Comparative and Physiological Psychology,* 1967, **63,** 258–269.

Gibson, E. J., and Yonas, A. A developmental study of visual search behavior. *Perception and Psychophysics,* 1966, **1,** 169–171.

Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1956.

Hardy, L. H., Rand, C., and Rittler, M. C. H.R.R. pseudoisochromatic plates. *American Journal of Optometry,* 1954, **44,** 509–516.

Hill, S. D. The performance of young children on three discrimination-learning tasks. *Child Development,* 1965, **36,** 425–436.

Michels, K. M., and Zusne, L. Metrics of visual form. *Psychological Bulletin,* 1965, **63,** 74–86.

Munsinger, H., and Kessen, W. Structure, variability and development. *Journal of Experimental Child Psychology,* 1966, **4,** 20–49.

Munsinger, H., Kessen, W., and Kessen, M. L. Age and uncertainty: Developmental variation in preference for variability. *Journal of Experimental Child Psychology,* 1964, **1,** 1–15.

Osler, S. F., and Kofsky, E. Stimulus uncertainty as a variable in the development of conceptual ability. *Journal of Experimental Child Psychology,* 1965, **2,** 264–279.

Piaget, J. *The origin of intelligence in the child.* New York: Harper, 1952.

Podell, J. E. Perception of mirror images and learning to discriminate them. Paper given at Midwestern Psychological Association, March 1965.

Pollack, R. H. Contour detectibility thresholds as a function of chronological age. *Perceptual Motor Skills,* 1963, **17,** 411–417.

Ricciuti, H. N. Geometric form and detail as determinants of comparative similarity judgments in young children. In *A basic research program on reading.* U. S. Office of Education Coop. Res. Project #639, 1963.

Rudel, R. G., and Teuber, H. L. Discrimination of direction of line in children. *Journal of Comparative and Physiological Psychology,* 1963, **56,** 892–898.

Sekuler, R. W., and Rosenblith, J. F. Discrimination of direction of line and the effect of stimulus alignment. *Psychonomic Science,* 1964, **1,** 143–144.

Stern, W. *Psychology of early childhood.* New York: Holt, 1924.

SUCHMAN, R. G. New look at young children's perceptual skills. Paper presented at the 75th annual convention of the American Psychological Association, 1967.

SUCHMAN, R. G., AND TRASBASSO, T. Stimulus preference and cue function in young children's concept attainment. *Journal of Experimental Child Psychology,* 1966, 3, 188–198.

THOMAS, H. Preference for random shapes: Ages six through nineteen years. *Child Development,* 1966, **37,** 843–859.

WELCH, L. The span of generalization below the two-year age level. *Journal of Genetic Psychology,* 1939, **55,** 269–297.

WERNER, H. *Comparative psychology of mental development.* New York: International Universities Press, 1957.

WESTCOTT, M. R., AND TOLCHIN, M. Studies of perceptual inference in children and adults. Paper given at Eastern Psychological Association, April, 1966.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 432–442 (1969)

Digit Span, Practice and Dichotic Listening Performance in the Mentally Retarded[1,2]

RICHARD C. URBANO[3] AND KEITH G. SCOTT
University of Illinois

The role of digit span and practice on the dichotic listening performance (DLP) of mentally retarded children was assessed in two experiments. In Exp. I, the DLP of two groups of 10 *Ss* was compared in a 2 × 5 × 2 × 3 factorial design—Groups (high digit span, 6; low digit span, 4) × Series length (1, 2, 3, 4, 5 digits/ear) × Half span (1st, 2nd) × Practice (Day 1, 2, 3). Digit span had a significant effect on DLP while practice did not. Experiment II attempted to replicate some of the effects of Exp. I and evaluate the role of maturational level in DLP with digit span held constant. The DLP of two groups of 10 *Ss* with like mean digit spans (4.3 and 4.5) but different MAs (10-0 and 6-3) was compared using the factorial design of Exp. I. The results of Exp. I were replicated and no significant effect of MA was found.

Attention and memory have been of central concern in research and theory on learning in the retarded (Scott and Scott, 1968; Zeaman and House, 1967). One task that involves both processes is the dichotic listening task introduced by Broadbent to investigate his information processing model (Broadbent, 1958). In the work of Inglis (1965) and Neufeldt (1966) the dichotic listening task and Broadbent's theory appear as important in assessing developmental and capacity differences.

Broadbent's model includes an attentional component in the perceptual (P) system, and a memory component in the storage (S) system. The P system is a limited capacity processor that handles inputs successively. In contrast, the S system is a large capacity processor that stores inputs when the P system is filled. In the dichotic listening task the *S* is pre-

[1] This investigation was supported by Grant MH-07346 from the National Institute of Mental Health and Grant HD 02898 from the National Institute of Child Health and Human Development both of the United States Public Health Service. The first study was submitted by the first author in partial fulfillment of the MA requirements at the University of Illinois.

[2] The authors express their appreciation to Dr. Louis Belinson, Superintendent and Mr. William Chambers, Chief Psychologist, Lincoln State School, for their cooperation in conducting the studies.

[3] Now at the University of Washington.

sented with pairs of digits simultaneously, one to each ear. At rates of about two pairs per second the subject reports the digits heard in one ear (the first half span) and then in the other (the second half span). The first half span is seen as a measure of the P system and the second as a measure of the S system.

Inglis, Sykes, and Ankus (1968) showed developmental effects on performance with the task both in children and in the aged; a result that was extended to cover capacity differences by Neufeldt (1966).

Neufeldt compared the performance of two retarded groups, organic (O) and familial (F), to the performance of one normal group matched or CA with the retardates (NCA) and a second normal group matched on MA with the retardates (NMA). Following Inglis's procedure, Neufeldt tried to match the groups on digit span. The mean digit spans of the O, F, and NMA groups were essentially the same (O = 4.53, F = 4.67, and NMA = 4.73), while that of the NCA group was significantly higher (NCA = 5.60). It is also interesting to note that the mean MA of the NCA group was higher than that of the other groups (O = F = NMA = 9.25; NCA = 14.39).

The performance of the NCA group was found to be superior to the performance of the other three groups which performed at essentially the same level. The results were interpreted to indicate that learning and maturational advantages of the NCA group were responsible for its superior performance. The fact that the NCA group differed from the other groups in digit span and MA leads to several alternative explanations of the data. The differences observed could be a function of: (1) the NCA's superior digit span; (2) the NCA's superior MA; or (3) a joint function of both superior MA and digit span.

On the basis of a series of experiments by Dodwell (1964) which reveal no practice effect, Neufeldt used the same groups of *Ss* in successive experiments. Dodwell used a practice list of 10 dichotic items prior to the beginning of each experimental session. It is possible that the long practice list eliminated any practice effect. Since Neufeldt used a short practice list, Dodwell's findings may not be applicable.

The present studies then, investigate the effects of digit span and practice on dichotic listening performance in a retarded population to further examine Neufeldt's interpretations and findings.

EXPERIMENT I

Subjects. The *Ss* were from the ambulatory, educable, and trainable residents at the Lincoln State School, Lincoln, Illinois. On the basis of performance on a tape recorded, modified WAIS digit span forward test, *Ss* were placed in the high digit span (HDS) and low digit span (LDS)

groups. Performance on the Peabody Picture Vocabulary Test was used to determine the MA and IQ of each *S*. The 10 *S*s in the HDS group (mean MA = 9-1, mean CA = 14-9, mean IQ = 68) had a mean digit span of 6.2 while the 10 *S*s in the LDS group (mean MA = 8-7, mean CA = 14-8, mean IQ = 71.8) had a mean digit span of 4.0.

Apparatus. A Roberts 720A tape recorder was used to present the stimuli through a pair of Grason-Stadler ANB-H-1 headphones. Each earphone was connected separately to one channel of the tape recorder. The sets of digits presented were taken from Neufeldt (1966) and recorded on each channel as shown in Table 1. A master tape was made with each of the nine digits, 1–9, recorded at the same intensity level, ±1 db, and less than 500 msec in duration. Copies of the master tape were made. Each digit was cut out of the master tapes and spliced together to make the two channels. The channels thus fabricated were transferred to the two channels of the stereo tape. Simultaneity of each

TABLE 1
SETS OF DIGITS TAKEN FROM NEUFELDT (1966) AND RECORDED ON EACH CHANNEL

	Channel 1	Channel 2
Practice series		
A	3	Blank
B	Blank	7
C	3	7
Test series	5	8
	7	6
	4	1
	6	3
	39	72
	85	17
	38	59
	65	28
	592	174
	793	462
	479	836
	584	719
	5638	2941
	9754	8362
	6542	7918
	9356	4271
	81324	96571
	74682	31579
	57841	29356
	38671	15492

pair was checked using a graphic level recorder. The rate of presentation was two pairs per second with a 10-second recall interval between sets.

Three daily sessions of dichotic listening were administered to each *S*. Each session with a given *S* began with a brief familiarization with the headphones and the presentation of the following instructions (the instructions were taken from Neufeldt, 1966, pp. 6 and 7): "Now listen carefully. You are going to hear a number. I want you to tell me what number you hear." Practice Series A was played. If *S* repeated the correct number, the procedure was repeated with Practice Series B. If *S* did not respond correctly, the volume was increased and Practice Series A repeated. *S* was then told: "Now you are going to hear two numbers together, one from each ear. Tell me what you hear." Then Practice Series C was presented. The Test Series was begun. The practice series was used to minimize differences in acuity and allow the *S* to adapt to the experimental procedure. Prior to each change in length the *S* was informed: "Now you are going to hear (N) numbers, (N/2) in each ear. Tell me what you hear."

Using the presentation order of series length 1 through 5 for all subjects confounds Series length and within session effects. Without very extensive pretraining and with the difficulty of the longer series lengths, counterbalancing or randomization of presentation order to eliminate the confounding was rejected. Zeaman and House (1960) have demonstrated that retarded children receiving a difficult experimental condition first quickly form a failure set, thus, producing a worse confounding. In addition, in a recently completed study Urbano and Scott (1969) found no significant within session order effects.

The scoring procedure was as follows: The first digit recalled was taken to indicate which Half span was being attempted first. Each correctly recalled digit was counted, order considered. The total number of digits correctly recalled from each Half span for each Series length was used to calculate the proportion correct for that Half span and Series length. Sets where the first digit recalled was not the first digit presented on one of the Half spans were given a score of 0. The latter scoring procedure assumes that the *S* will recall the digits from one Half span before recalling any from the other Half span. Although other methods of scoring can be used, Neufeldt (1966) and Inglis and Tansey (1967) have shown that while more lenient scoring methods do yield systematically higher scores, they do not produce significant interactions with capacity or age differences.

Design. A factorial design was used with one between *Ss* factor, Groups (HDS and LDS) and three within *Ss* factors, Series length (1, 2, 3, 4,

and 5 digit/ear), Half span (first and second), and Practice (Days 1, 2, and 3).

Results

A $2 \times 5 \times 2 \times 3$ (Groups $\times$ Series length $\times$ Half span $\times$ Practice) factorial analysis of variance was performed on the proportion correct recalls. *Ss*, the replication factor, was a random factor, while the other factors were fixed. An analysis of variance was also performed on the arcsin transformation of the proportions yielding essentially the same results and is not reported here. Figure 1 presents the mean proportion

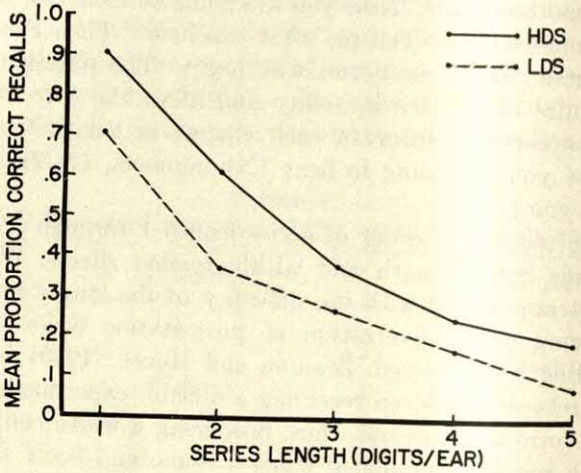


FIG. 1. Proportion correct recalls as a function of Series length is presented for the HDS and LDS groups.

correct recalls for the HDS and LDS groups, collapsed across Half span and Practice levels, as a function of Series length. The main effect of Groups, F (1, 18) = 22.35, $p < .001$, results from the superior performance of the HDS group at every Series length. The main effect of Series length, F (4, 72) = 159.29, $p < .001$, is a result of the dramatic decrease in the performance of both groups from the shortest to the longest Series length. There is no significant interaction between Groups and Series length, F (4, 72) = 2.50, n.s.

The effect of Half span can be seen in Figure 2 which presents the mean proportion correct recalls for the HDS and LDS groups, collapsed across Practice levels, as a function of Series length and Half span. The superior performance of both groups on the first Half span in comparison to that on the second Half span is reflected in the significant main effect of Half span, F (1, 18) = 139.68, $p < .001$, and the significant Series length by Half span interaction, F (4, 72) = 10.58, $p < .001$. First Half

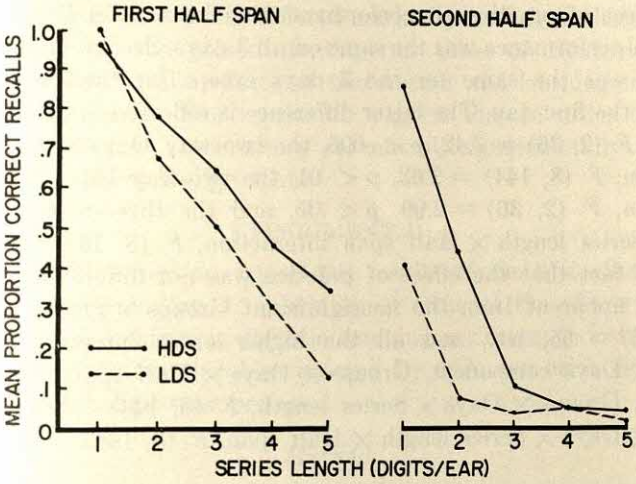


FIG. 2. Proportion correct recalls for the first and second Half span as a function of Series length is presented for the HDS and LDS groups.

span performance is at the same high level for both groups for the shortest series length. The HDS group shows less of a decrement in performance as Series length increases. Second Half span performance of the HDS group is superior to that of the LDS group on the shortest Series length but not on the longer Series lengths. The latter differential relationship between groups as a function of Series length is reflected in the significant Groups $\times$ Series length $\times$ Half span interaction, F (4, 72) = 15.52, $p < .001$.

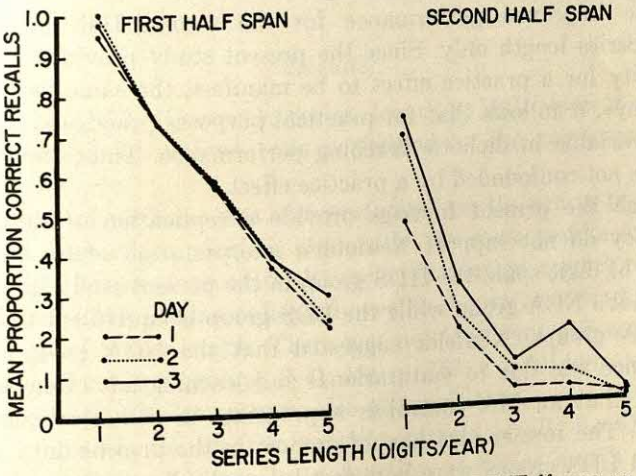


FIG. 3. Proportion correct recalls for the first and second Half span as a function of Series length is presented for Days 1, 2, and 3.

The effect of practice on performance can be seen in Figure 3. First Half span performance was the same on all 3 days. Second Half span performance was the same for the 3 days except for the shortest series length on the first day. The latter difference is reflected in the main effect of Days, F (2, 36) = 7.42, $p < .005$, the two-way Days × Series length interaction, F (8, 144) = 2.63, $p < .01$, the two-way Days × Half span interaction, F (2, 36) = 2.99, $p < .05$, and the three-way interaction, Days × Series length × Half span interaction, F (8, 144) = 2.33, $p < .025$. The fact that the effect of practice was not different for the two groups is apparent from the nonsignificant Groups × Days interaction, F (2, 36) = .55, n.s., and all the higher order interactions with a Groups × Days component, Groups × Days × Half span, F (2, 36) = 1.97, n.s., Groups × Days × Series length, F (8, 144) = 1.40, n.s., and Groups × Days × Series length × Half span, F (8, 144) = 1.11, n.s.

Discussion

The basic findings of the dichotic listening technique, superior first Half span performance and poorer performance as Series length increases (Broadbent, 1958; and Inglis, 1965) were replicated in the present study. In terms of Broadbent's model, the superior first Half span performance of the HDS group reflects a larger capacity P system. The latter also suggests that ordinary digit span provides a sensitive measure of P system capacity. With a larger capacity P system the HDS group had to store less information in the S system than did the LDS group. Thus, the HDS group's superior second Half span performance cannot be attributed to superior S system functioning.

Practice improved performance for the second Half span at the shortest Series length only. Since the present study provided a sufficient opportunity for a practice effect to be manifest, the same list on 3 successive days, it follows that for practical purposes practice is not an important variable in dichotic listening performance. Thus, Neufeldt's results were not confounded by a practice effect.

Although the present findings provide a replication of the Neufeldt study, they do not support Neufeldt's interpretation of his results. On the basis of digit span, the HDS group in the present study is equivalent to Neufeldt's NCA group while the LDS group is equivalent to the O, F, and NMA groups. Neufeldt suggested that the NCA group's superior performance was due to maturational and learning advantages. He also suggested that an MA control is appropriate in dichotic listening performance. The former idea has no support in the present data, since the HDS and LDS groups were both institutionalized retardates. The latter idea also suffers, since the HDS and LDS groups had mean MAs which

differed only by 6 months but differed markedly in dichotic listening performance. Neither the present finding nor those of Neufeldt provide a basis for assessing the effects of maturation and learning in dichotic listening performance. The results do suggest that in order to answer the latter question it is necessary to control, experimentally or statistically, for the effects of digit span.

EXPERIMENT II

Experiment II aimed to extend the results of the first study in two ways: first, by providing a replication of some of the effects observed there; second, by examining dichotic listening performance in two groups of different maturation level as indexed by MA, but with comparable performance on a digit span pretest.

Method

Subjects. The *Ss* were from the ambulatory, educable, and trainable residents at the Lincoln State School, Lincoln, Illinois. Each *S* was administered the Peabody Picture Vocabulary Test and a tape recorded, modified WAIS digit span test. Two groups of *Ss* were selected on the basis of MA. The high MA (HMA) group had a mean MA = 10-0, mean CA = 14-8, and mean IQ = 71.8 while the low MA (LMA) group had a mean MA = 6-3, mean CA = 13-4, and mean IQ = 58.0. After the sampling a check revealed that the mean digit spans of the two groups were similar (HMA, $\bar{x} = 4.3$, range 3–6, LMA, $\bar{x} = 4.5$, range = 3–5).

Procedure. Except for the substitution of MA for digit span as the basis for the Groups factor, the procedure and apparatus were the same as Exp. I.

Results

A $2 \times 5 \times 2 \times 3$ (Groups × Series length × Half span × Practice) factorial analysis of variance was performed on the proportion correct recalls.

An analysis of variance was also performed on the arcsin transformation of the proportions yielding essentially the same results and is not reported here. Figure 4 presents the mean proportion correct recalls for the HMA and LMA groups, collapsed over Half span and Practice levels, as a function of Series length. There was no significant difference in the performance of the two groups, F (1, 18) = .757, n.s. All two- and three-way interactions with Groups were also nonsignificant.

Figure 5 presents the mean proportion correct recalls for Days 1, 2, and 3, collapsed over Groups, as a function of Series length and Half span. Although performance was superior on the first Half span and the

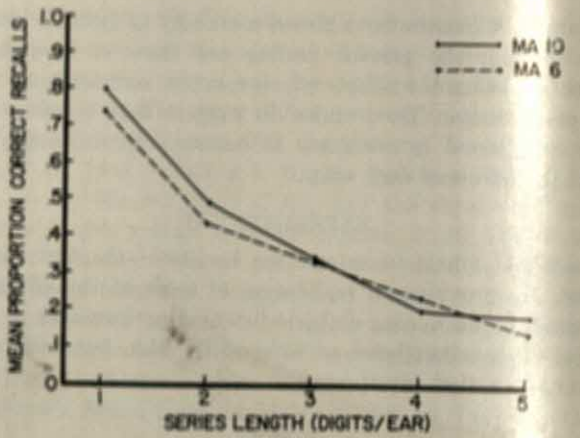


FIG. 4. Proportion correct recalls as a function of Series length is presented for the HMA and LMA groups.

shorter Series lengths as reflected in the significant main effects of Half span, F (1, 18) = 172.66, $p < .001$, and Series length, F (4, 72) = 99.53, $p < .001$, the decrement in performance with increased Series length was more drastic for the second Half span as reflected in the significant Series length × Half span interaction, F (4, 72) = 10.13, $p < .001$. Performance on the first Half span improved little over days while the performance on the second Half span improved on the shortest Series length from Day 1 to Day 2. The latter improvement in performance is reflected in

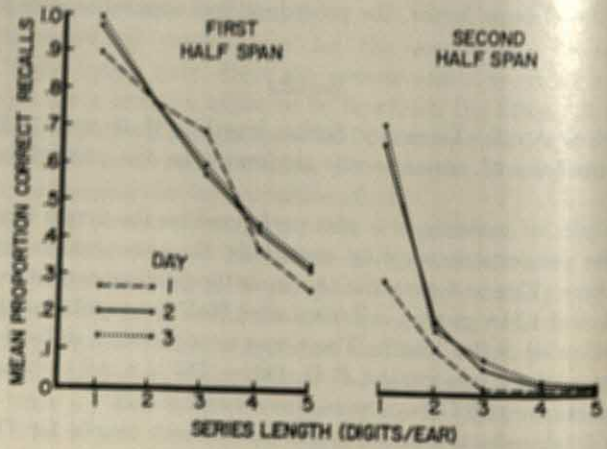


FIG. 5. Proportion correct recalls for the first and second Half span as a function of Series lengths is presented for Days 1, 2, and 3.

the main effect of Practice, F (2, 36) = 11.69, $p < .001$, the two-way interactions of Practice × Series length, F (8, 144) = 8.73, $p < .001$, and Practice × Half span, F (2, 36) = 7.16, $p < .005$, and the three-way interaction of Practice × Series length × Half span, F (8, 144) = 4.41, $p < .001$.

Discussion

The results of the present study provide further replication of the major task variables affecting dichotic listening performance in the mentally retarded. In addition, they show that groups of like digit span, but different in maturational level perform similarly. As a matter of practical concern, this suggests that digit span, and not MA, is the more appropriate basis for matching retarded and probably normal children in investigations of dichotic listening.

Theoretically, the ability of digit span to account for performance on a dichotic listening task is of some importance. The claimed significance of the dichotic listening task is that it provides separate measures of both memory and attentional processes. If performance on both Half spans can be mainly predicted from performance on a memory task (digit span) alone then this claim is weakened. The failure to find a significant main or interactive effect of MA with dichotic listening task variables constitutes such negative evidence. Further, since MA and digit span are highly correlated while dichotic listening performance is predicted mainly from digit span and not MA, the present data suggest that the dichotic listening task is measuring a parameter of individual variation not indexed by MA.

In view of the systematic changes in dichotic listening performance as a function of normal and abnormal development (Inglis, 1965) and the potential of indexing an ability separate from MA, further research is needed. Two interdependent lines of investigation, covering the structure of individual differences and the relationship among task variables, are necessary to explore the theoretical and diagnostic potential of the dichotic listening task. The nature and structure of individual differences can be described using the techniques of psychometric psychology to delineate dimensions of systematic variation between individuals. The techniques of experimental psychology provide for a mapping of the relationship among task variables, giving a framework of basic psychological processes on which individuals may vary. Thus, a thorough explanation of the place of memory and attention in development will be rooted to both experimental and psychometric psychology using what Cronbach (1957) has called a united methodology. For example, following this methodology, a direct extension of the present study should have

two related facets. First, the relationship of MA, digit span, and systematic difference in dichotic listening should be delineated using a population with a wide range of MA and digit span. Second, the relationship of systematic individual differences and parameters should be determined by evaluating the effects of manipulating such task parameters as presentation rate, recall interval, and stimulus material.

REFERENCES

BROADBENT, D. E. The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, 1954, **47**, 191–196.

BROADBENT, D. E. Successive responses to simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, 1956, **8**, 145–152.

BROADBENT, D. E. Immediate memory in simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, 1957, **9**, 1–11.

BROADBENT, D. E. *Perception and communication*. London: Macmillan (Pergamon), 1958.

CRONBACH, L. J. The two disciplines of scientific psychology. *American Psychologist*, 1957, **12**, 671–684.

DODWELL, P. C. Some factors affecting the hearing of words presented dichotically. *Canadian Journal of Psychology*, 1964, **18**, 72–91.

INGLIS, J. Immediate memory, age and brain function. In A. T. Welford and J. E. Birren (Eds.), *Behavior, aging, and the nervous system*. Springfield, Illinois: Thomas, 1965, pp. 88–113.

INGLIS, J., SYKES, D. K., AND ANKUS, M. Age differences in short-term memory. *Interdisciplinary Topics on Gerontology*, 1968, **1**, 18–43.

INGLIS, J., AND TANSEY, C. L. Age differences and scoring differences in dichotic listening performance. *Journal of Psychology*, 1967, **66**, 325–332.

NEUFELDT, A. H. Short-term memory in the mentally retarded: An application of the dichotic listening technique. *Psychological Monographs*, 1966, **80** (12, Whole No. 620).

SCOTT, K. G., AND SCOTT, M. S. Research and theory in short-term memory. In N. R. Ellis (Ed.), *International Review of Research in Mental Retardation*. Vol. 3, pp. 135–162. New York: Academic Press, 1968.

URBANO, R. C., AND SCOTT, K. G. Performance of retarded subjects on a two channel auditory—visual memory task. K. Scott (Ed.), Progress Report #1, Grant USPH HD 02898, University of Illinois, 1969.

ZEAMAN, D., AND HOUSE, B. J. Approach and avoidance in the discrimination learning of retardates. In B. J. House *et al. Learning and transfer in mental defectives*. Progress Report 2. NIMH USPHS, 1960, University of Connecticut, 32–70.

ZEAMAN, D., AND HOUSE, B. J. The relationship of IQ and learning. In R. M. Gagne (Ed.), *Learning and individual differences*. Columbus, Ohio: Merrill, 1967. pp. 192–212.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 443–460 (1969)

The Development of Cooperation in Alternative Task Situations[1]

DENNIS E. MITHAUG
University of Washington

This study is a follow-up of previous work by Mithaug and Burgess (1967, 1968) which suggested that when (*a*) subjects are permitted to watch their points accumulate on individual and group task counters and (*b*) are free to work together on the group task or alone on individual tasks, i.e., during the condition of group and individual reinforcement with individual feedback, cooperation does not develop. The present study investigates this finding in more detail. Children, ranging from five to ten years of age, were divided into four three-member groups. The data indicate that cooperation can be produced in alternative task situations when (*a*) rewards resulting from the group task are greater than rewards resulting from individual tasks and (*b*) task situations permit subjects to learn to discriminate between group and individual tasks in terms of their relative payoffs.

An experimental analysis of developing cooperative patterns in two-person groups has been applied by a number of investigators (Azrin & Lindsley, 1956; Cohen, 1962; Cohen & Lindsley, 1965; Lindsley, 1966). Mithaug and Burgess (1967, 1968) have extended this method of investigation to three-person groups. Groups were confronted with the task of accumulating points on electromechanical counters by using 14-key piano keyboard instruments. However, they were given no specific instructions for using the instruments to operate the counters. The only instructions given were that points accumulated on counters would be exchanged for pennies and that they could press any one key as fast and as many times as desired during a given trial period.

The question under investigation was: What conditions will produce cooperation? Although not explicitly defined or discussed in these terms, their working definition of cooperation was coordinated behavior that completes a common task, i.e., produces points on a group counter.

The tasks involved were of two types, one group task and three identical individual tasks. For the group task, represented by a group counter situated in front of the subjects, points could be earned jointly when all three pressed a predesignated key at the same time. For individual

[1] For reprints, write to the Experimental Education Unit, Child Development and Mental Retardation Center, University of Washington, Seattle, Washington 98105.

tasks, represented by three counters situated beside the instruments of each subject, points could be earned separately by pressing the correct key.

Mithaug and Burgess (1968) observed that a coordinated pattern involving a division of labor, two persons holding the predesignated key and one person pressing it, was more effective at producing points on the group counter than pressers chanting "one, two, three, four," together in order to aid coordination, which in turn was more effective than three pressers responding rapidly in an uncoordinated fashion.[2]

Coordinated behavior resulting in increased point accumulation rates occurred during a condition of group reinforcement with individual feedback. However, in alternative task situations, cooperation did not develop. During this condition, subjects could choose to work on group or individual tasks; both group and individual counters were backed by penny rewards. The data indicated that point rates for individual counters were higher than for the group counter.

Mithaug and Burgess suggested that making the interdependent task more rewarding than the independent tasks may motivate subjects to switch tasks and work on the group counter rather than the individual counters. The purpose of the present study is to investigate the questions of whether cooperation can be produced during group and individual reinforcement with individual feedback when completion of an interdependent task results in greater rewards than completion of independent tasks.

METHOD

Group composition. The subjects participating in this study were normal children ranging in age from 5 to 10 years. They were divided into three-member groups comprising various combinations of sex. In all, four separate groups participated. Group 1 consisted of 3 boys, two friends and a stranger, ages ranging from 8 to 10. Group 2 consisted of three girls, all sisters, ages ranging from 5 to 10. Group 3 consisted of 2 girls and a boy, all siblings, ages ranging from 6 to 10. The fourth group consisted of 2 girls and a boy, all strangers, ages ranging from 8 to 10 years.

Apparatus. The apparatus consisted of three 14-key piano keyboard

[2] In the previous work by Mithaug and Burgess, the term "group response" was used to identify the accumulation of points on the group counter. The term is somewhat misleading because it confuses two important and distinct events, the accumulation of points on the group counter *and* the particular behavior pattern developed by the group to produce points. For this reason, the term task or task completion rate will be used here to identify point rates on the counters.

instruments wired to a counting apparatus on the experimenter's desk. On top of each instrument were two readily visible lights, one on the left and one on the right. Electromechanical counters on the experimenter's desk recorded the number of times a key representing a correct response was depressed. Similarly, electromechanical counters were situated beside each instrument to record the correct key presses. The three keyboard instruments were aligned against a wall in a laboratory room. A 12 × 36-inch screen, on which was a series of lights, was hung in the center of this same wall. This light screen was also wired to the apparatus on the experimenter's desk. The experimenter, by activating a switch on his desk, illuminated a light on the light screen corresponding to a particular key on the subjects' keyboard instruments. Another series of switches on the desk enabled the experimenter to control the lights on the three keyboard instruments. If the enabling switches were placed at the "on" position, and if the subjects depressed the key corresponding to the illuminated light on the light screen, lights on their instruments were illuminated.

Procedure. The general procedure was as follows. The experimenter activated a switch controlling a particular light on the light screen. The subjects pressed the key of their choice. If the chosen key corresponded to the light illuminated on the light screen, and if the enabling switches on the experimenter's desk were at the "on" position, lights on the subjects' instruments flashed on. The electromechanical counters at the experimenter's desk and the counters situated beside the subjects' keyboard instruments tabulated points for correct key presses. The experimenter could make these latter counters inoperative by activating appropriate switches at his desk.

Finally, the single light and electromechanical counter in the center of the light screen were activated: (*a*) if the controlling switches at the experimenter's desk were in the "on" position; (*b*) if all three subjects pressed the key corresponding to the illuminated light on the light screen; and (*c*) if each subject pressed the correct key at the same time, within .5 second of each other. When all three subjects pressed the correct keys within this time interval, a group point was recorded by the group counter on the light screen, by a counter situated on the experimenter's desk, and by a Gebrands' Cumulative Recorder also situated at the desk.

Experiment I

Procedure. As noted in the introduction, Mithaug and Burgess (1968) were unable to produce cooperation during group and individual reinforcement with individual feedback. During this condition 100 points were exchanged for one cent for both group and individual counters.

The hypothesis is: When greater rewards for achieving the group rather than individual tasks are provided subjects during this condition, cooperation will emerge. Experiment I pretests this notion, using one of the same groups that participated in the previous Mithaug and Burgess study (1968).

The three subjects were instructed that their task was to accumulate points on the counters and that, to achieve it, they could press any single key on their keyboard instruments as fast and as many times as desired during a given trial. When the group counter reached 100 points, each received one cent. When both lights flashed on the keyboard instruments, 100 points accumulated on an individual counter returned one cent to that subject. When one light flashed on the instrument, 1000 points returned one cent.

The session began with a light on the light screen, indicating to the subjects they could begin pressing. The response time was divided into two-minute intervals. A clock wired to the apparatus stopped when the light flashed off the screen, and the apparatus was turned off.

Group and individual counters were in operation tabulating points for simultaneous key presses and correct individual key presses, and the lights flashed to provide feedback for these correct key presses. The session began with Condition A, two lights flashing, indicating the exchange rate for the individual counters was 100 points = one cent. Condition B followed with one light flashing, indicating to the subjects that the exchange rate for the individual counters was 1000 points = one cent. The experiment ended with Condition A, when both lights flashed and 100 points were again exchanged for one cent on individual counters. The group counter returned one cent for 100 points for the entire experiment.

Results. There appeared to be little difference in the rates between conditions. The task rate for individual counters was higher than the task rate for the group counter during both A and B conditions. Also, no interaction involving coordinated key presses, e.g., two holders and a presser, was observed.

Discussion. There seems to be no support for the hypothesis that greater rewards for achieving the group task will produce cooperation. However, the subjects appeared to be trying to see who could accumulate the most points. This behavior pattern may have overshadowed any possible effects of the ratio between counters.

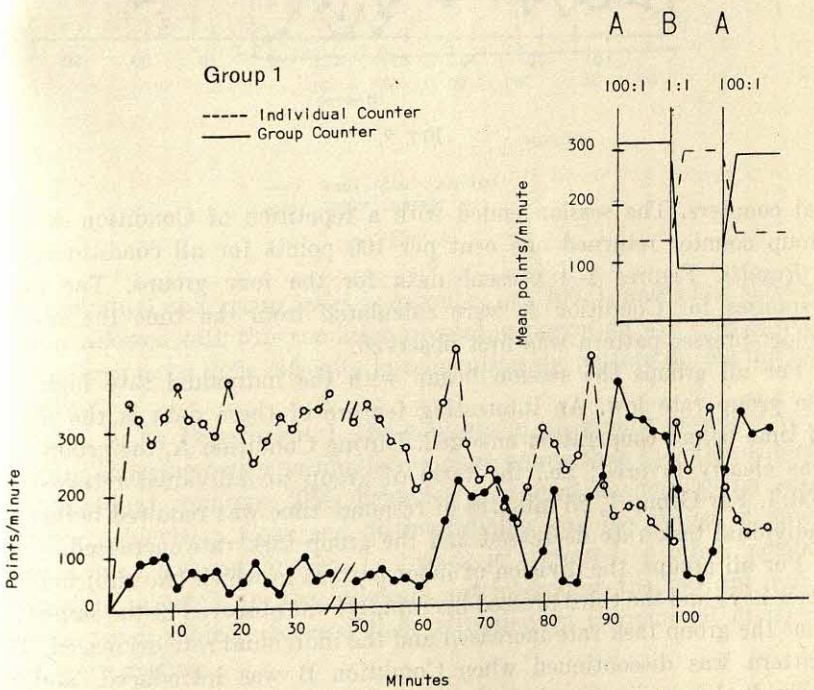
Experiment II

Experiment II investigates the possibility that a more dramatic difference in exchange rates between group and individual tasks may be necessary to get cooperative patterns started. Furthermore, a longer time

may be required before subjects recognize the advantages of working on the interdependent task.

Procedure. Group 1 and three additional groups participated in this experiment. Each group contained three subjects, ranging in age from five to ten years. Subjects were given the same instructions as in Experiment I. When the group counter reached 100 points, each received one cent. When both lights flashed on the instruments, 100 points accumulated on individual counters returned one cent. However, when only one light flashed on the individual counter, 10,000 points returned one cent.

The session began with Condition A, one instrument light flashing, indicating the exchange rate for individual counters was 10,000 points for one cent. This condition was prolonged, giving ample time for the development of coordinated key presses, along with an increase in the group rate. Condition B followed with two instrument lights flashing, indicating the exchange rate was now 100 points for one cent on individ-



FIGS. 1–4. The lower graph presents the points per minute and the upper graph the means of these rates for Conditions A, B, A. The ratios (100:1, 1:1) represent the proportion of point-reward return for the individual counters to point-reward return for the group counter.

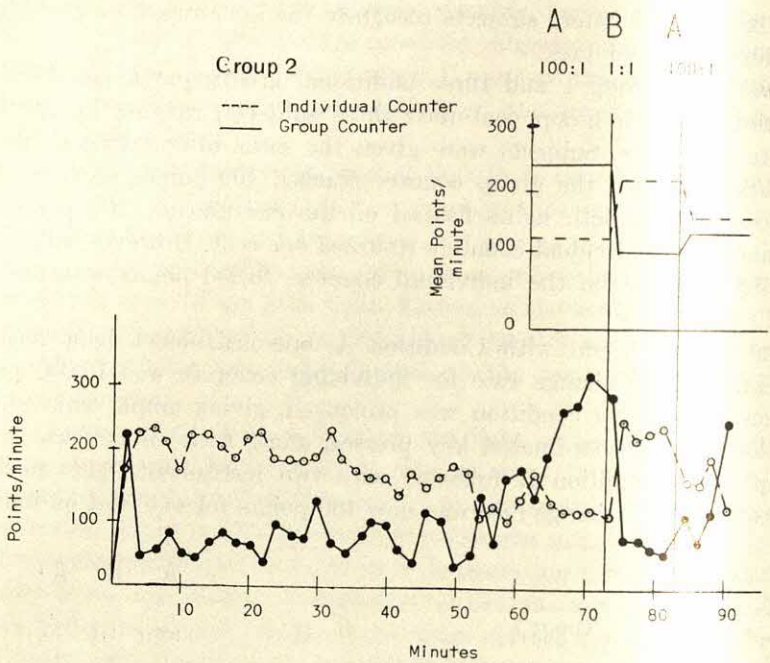


FIG. ?.

ual counters. The session ended with a repetition of Condition A. The group counter returned one cent per 100 points for all conditions.

Results. Figures 1–4 present data for the four groups. The mean responses for Condition A were calculated from the time the holder-holder-presser pattern was first observed.

For all groups the session began with the individual rate high and the group rate low. An interesting feature of these data is the length of time before cooperation emerged. During Condition A, the group task was clearly favored; i.e., the ratio of group to individual returns was 100:1. For Group 1, 85 minutes of response time was required before the individual task rate decreased and the group task rate increased.

For all groups, the division of labor pattern in which two subjects held their keys and the third pressed his rapidly, was observed at the same time that the group task rate increased and the individual rate decreased. This pattern was discontinued when Condition B was introduced, and the individual rate increased and the group rate decreased for all groups. The pattern was again observed during the final Condition A when the group rate increased and the individual rate decreased.

Discussion. These data suggest that when the ratio of exchange rates

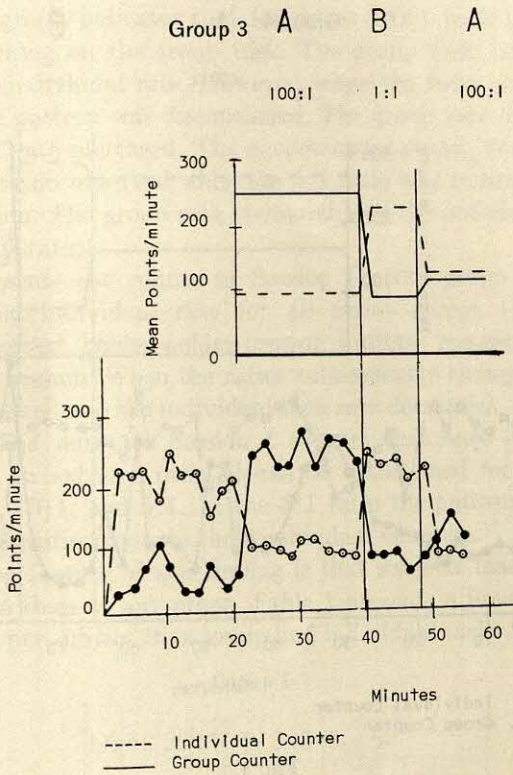


FIG. 3.

for individual and group tasks is extreme, i.e., 100:1, and when sufficient time is allowed with this condition present, cooperation will emerge. Also, there seems to be little difficulty in reinstating the pattern during the final condition.

Again, as in Experiment I, the subjects in all groups appeared to be initially concerned with accumulating more points than the other subjects on individual counters. This behavior continued as long as no single member achieved a clear and distinguishable lead over other members. Once a lead was established, i.e., when it appeared to other members impossible to catch up, the followers lost interest in the race for points. At this time, subjects focused their attention on the amount of money they were getting.

Experiment III

Procedures. The results of Experiment II raise interesting questions concerning emerging cooperative patterns under less extreme conditions

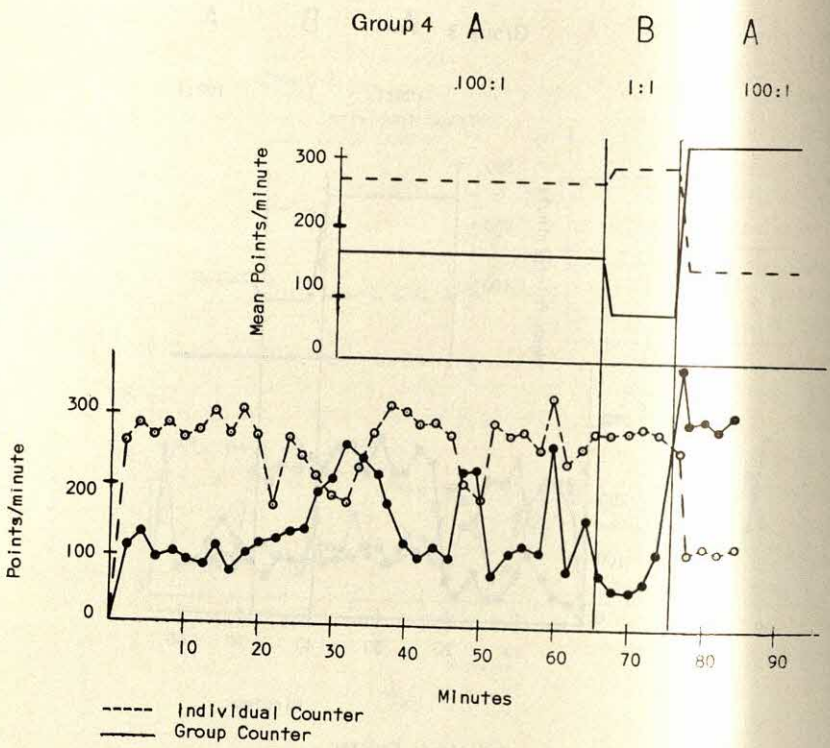


FIG. 4.

of differential reward. Therefore, using Group-1 subjects, this experiment investigates reward ratios of 100:1, 50:1, 25:1, 10:1, 5:1, and 3:1. The first number represents the point-reward returned for the individual counter and the second the point-reward return for the group counter for Session 1. The sequence of presentation was as follows: 100:1, 50:1, 25:1, 10:1, 5:1, 3:1, 5:1. The only change in procedure was an alteration in the method of presenting the ratios. Instead of being instructed that one light represented one ratio and two lights another, subjects were given exchange booklets. On each page of the booklet was a different rate for individual counters; for the group counter, one point always returned one cent. Subjects were not permitted to leaf through the booklet. They proceeded from front to back, turning one page at a time upon instruction from the experimenter.

For Session 2, the exchange booklets were replaced with booklets having the following sequence of ratios: 100:1, 50:1, 25:1, 10:1, 5:1, 3:1, 1:1, 3:1, 5:1, 10:1.

Results. The data for Session I of this experiment are presented in

Figure 5. The graph indicates that for ratios 100:1 to 5:1, the subjects cooperated working on the group task. The group task rate was higher than the mean individual rate. However, when the ratio changed to 3:1, the cooperative pattern was discontinued. The group rate dropped as the individual task rate increased. The cooperative pattern, two holders and a presser, was again observed when the 5:1 ratio was reintroduced at the end of the session. The group rate increased, and the individual task rate decreased at this ratio.

Figure 6 presents the results of Session II. The group rate remains higher than the individual rate for all ratios except 1:1. Here the previously observed holder-holder-pressor pattern re-emerged and remained for the session. When the ratios subsequently changed to 3:1, the group rate increased and the individual task rate decreased.

Discussion. The data for Session 1 suggest that once cooperation is established for an extreme ratio, it can be maintained for lesser ratios, i.e., 50:1, 25:1, 10:1, and 5:1. At the 3:1 ratio the pattern broke down, with subjects attending more to their individual counters.

An interesting feature of this finding is that subjects made less money at the 3:1 ratio than at any other. Table 1 presents a breakdown of the amount earned per subject from group and individual tasks for each ratio.

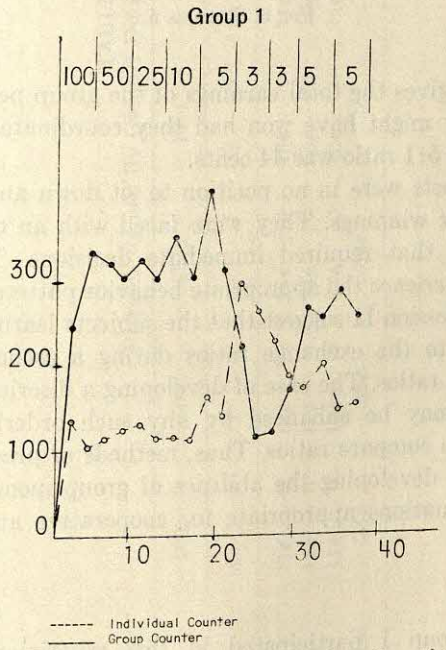


FIG. 5. The numbers heading the graph represent ratios of point-reward return for the individual counters to point-reward return for the group counter.

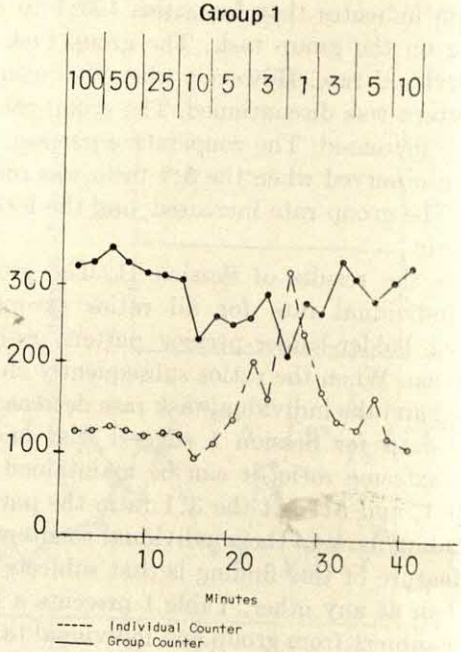


FIG. 6. Same as 5.

The bottom row gives the total earnings of the group per exchange ratio. The amount they might have won had they coordinated their efforts as they did during a 5:1 ratio was 44 cents.

However, subjects were in no position to sit down and calculate ways to maximize their winnings. They were faced with an ongoing problem-solving situation that required immediate decisions. They could only *learn through experience* the appropriate behavior patterns.

Results from Session II suggest that the subjects learn to respond more discriminatively to the exchange ratios during a second session of experience with the ratios. The ease of developing a discriminative response to these ratios may be enhanced by any such orderly sequence that allows subjects to compare ratios. Thus, methods of presenting the ratios are important in developing the abilities of group members to discriminate between situations appropriate for cooperative and individualistic behaviors.

Experiment IV

Procedure. Group I participated in this experiment in which the previous ratios were now presented randomly. All procedures remained

TABLE 1
Rewards Earned By Group 1 During Session 1 of Experiment III

	100:1		50:1		25:1		10:1		5:1		3:1		3:1		5:1		5:1	
	G	I	G	I	G	I	G	I	G	I	G	I	G	I	G	I	G	I
S-1	11¢	0	12¢	0	13¢	0	13¢	0	14¢	0	7¢	3¢	6¢	1¢	10¢	1¢	11¢	0¢
S-2	11¢	0	12¢	0	13¢	0	13¢	0	14¢	0	7¢	3¢	6¢	2¢	10¢	1¢	11¢	0¢
S-3	11¢	0	12¢	0	13¢	0	13¢	0	14¢	2¢	7¢	4¢	6¢	3¢	10¢	2¢	11¢	2¢
Total	33¢		36¢		39¢		39¢		44¢		31¢		24¢		34¢		35¢	

the same as in Experiment III except that the pages of the booklets were ordered from front to back in a pre-programmed random sequence.

Results. Figure 7 shows that the group rate is higher than the mean individual task rates for all ratios except 3:1 and 1:1, when the holder-holder-presser pattern was discontinued and subjects attended to their individual counters.

Discussion. The data from Experiment IV suggest that the method of presenting ratios produces different results. When ratios were presented in a decreasing sequence for the second time (Experiment III), subjects cooperated at all ratios except 1:1. However, with random presentations, subjects cooperated at all ratios except two of the three 3:1 ratios and three of the three 1:1 ratios. These findings suggest that subjects discriminate more when ratios are presented in a decreasing order than when they are presented randomly. Additional data are needed to support this notion, as the previous experimental history of Group 1 may account

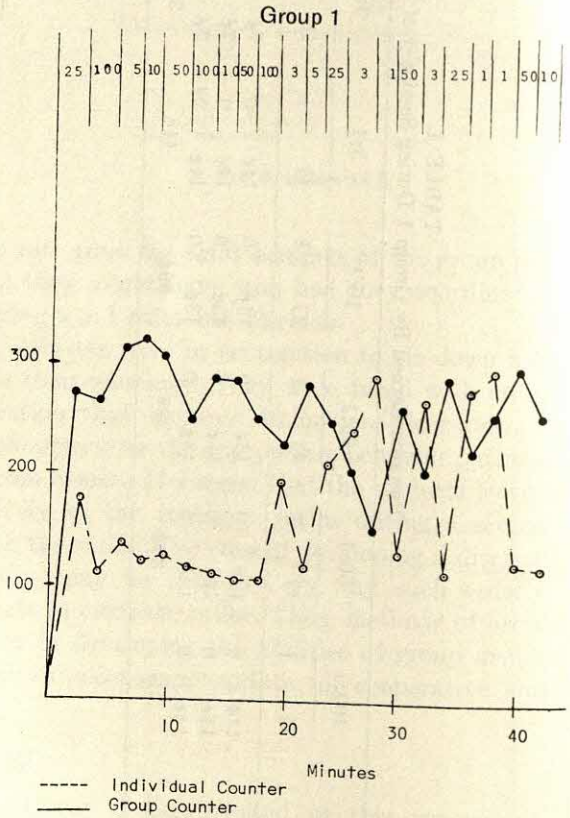


FIG. 7. Same as 5.

for the increased discrimination between ratios. The subjects of Group I may have learned discriminative response patterns of working together for some ratios and individually for others during the previous two experiments.

Experiment V

Experiment V investigates the possibility that the response patterns emerging when ratios were presented randomly during Experiment IV may have been different if subjects had not learned to discriminate between ratios, cooperating only when the group task led to greater returns than did the individual tasks.

Procedure. The same procedures were used for this experiment as for Experiment IV, but the subjects were from Group 4 instead of Group 1. Group 4's previous experimental history included Experiment II, but not Experiments III and IV. Therefore, Group 4 had experienced only

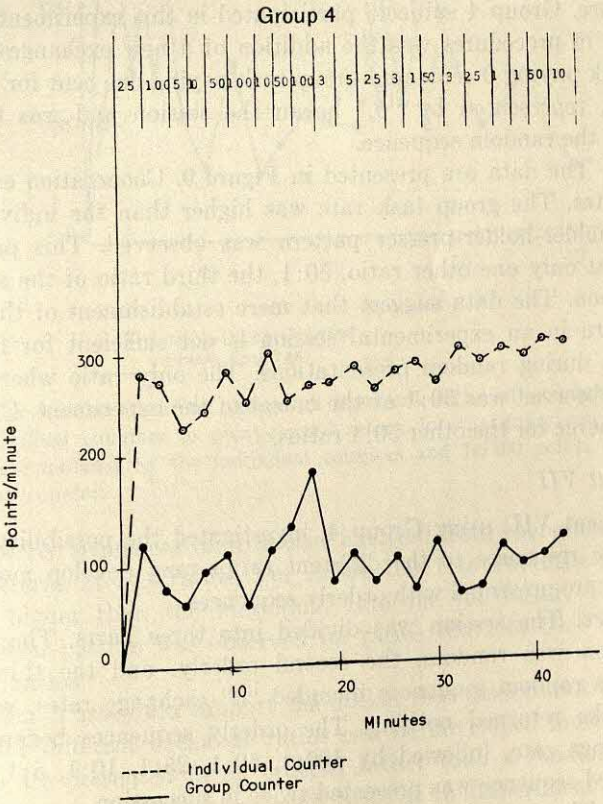


FIG. 8. Same as 5.

two exchange ratios, 100:1, for which a cooperative pattern developed, and 1:1, for which no cooperation emerged.

Results. Figure 8 presents data for this experiment. For all ratios, the mean individual task rate was higher than the group task rate. The holder-holder-presser pattern did not emerge.

The mean individual task rate was higher than the group rate for all ratios, even the most extreme 100:1, where it was clearly advantageous for subjects to cooperate as they had done for that ratio in Experiment II.

Experiment VI

Experiment VI investigates the possibility that a new exchange rate, individual counters operating but returning 0 cents, and the group counter returning one cent for 100 points, may start the subjects cooperating in the session. Once established, the pattern may generalize to other ratios as well.

Procedure. Group 4 subjects participated in this experiment. The only alteration in procedures was the addition of a new exchange rate: individual task paying 0 cents and group task paying one cent for 100 points. This rate, represented by "0," began the session and was then interspersed in the random sequence.

Results. The data are presented in Figure 9. Cooperation emerged for all "0" rates. The group task rate was higher than the individual rate, and the holder-holder-presser pattern was observed. This pattern was observed at only one other ratio, 50:1, the third ratio of the session.

Discussion. The data suggest that mere establishment of the cooperative pattern in an experimental session is not sufficient for further development during random presentations. The only ratio where cooperation was observed was 50:1 at the outset of the experiment. Cooperation did not emerge for the other 50:1 ratios.

Experiment VII

Experiment VII, using Group 4, investigated the possibility that discriminative responses to the different ratios may develop more rapidly if subjects are presented with orderly sequences.

Procedure. The session was divided into three parts. The first ratio presentation was random, the second orderly, and the third random again. The random sequences included "0" exchange rates, where individual tasks returned nothing. The orderly sequences began with the "0" exchange rate, followed by 100:1, 50:1, 25:1, 10:1, 5:1, 3:1, 1:1. The ordered sequence was presented twice in succession.

Results. The data for this experiment are presented in Figure 10. Dur-

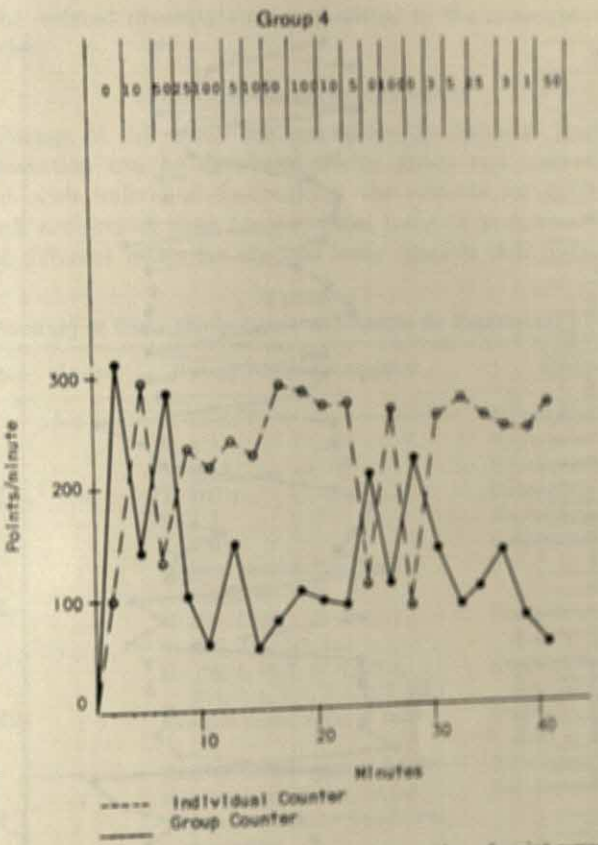


FIG. 9. The numbers heading the graph represent ratios of point-reward return for the individual counters to point-reward return for the group counter: 0 = 0¢ for points accumulated on the individual counters and 1¢/100 points accumulated on the group counter.

ing the orderly sequences the response changed from cooperation to individual patterns at 5:1 ratios. For ratios 100:1 to 10:1 the group task rate was higher than the individual rate for the latter. The holder-holder-presser pattern was observed for ratios 100:1 to 10:1 but not for 5:1 to 1:1 ratios.

Discussion. These data support the notion that discriminative response patterns for different exchange ratios can be developed more rapidly if the ratios are presented in an ordered rather than a random sequence. When they were presented randomly first, as in Experiment V and VI, no cooperation emerged. However, after some discrimination developed

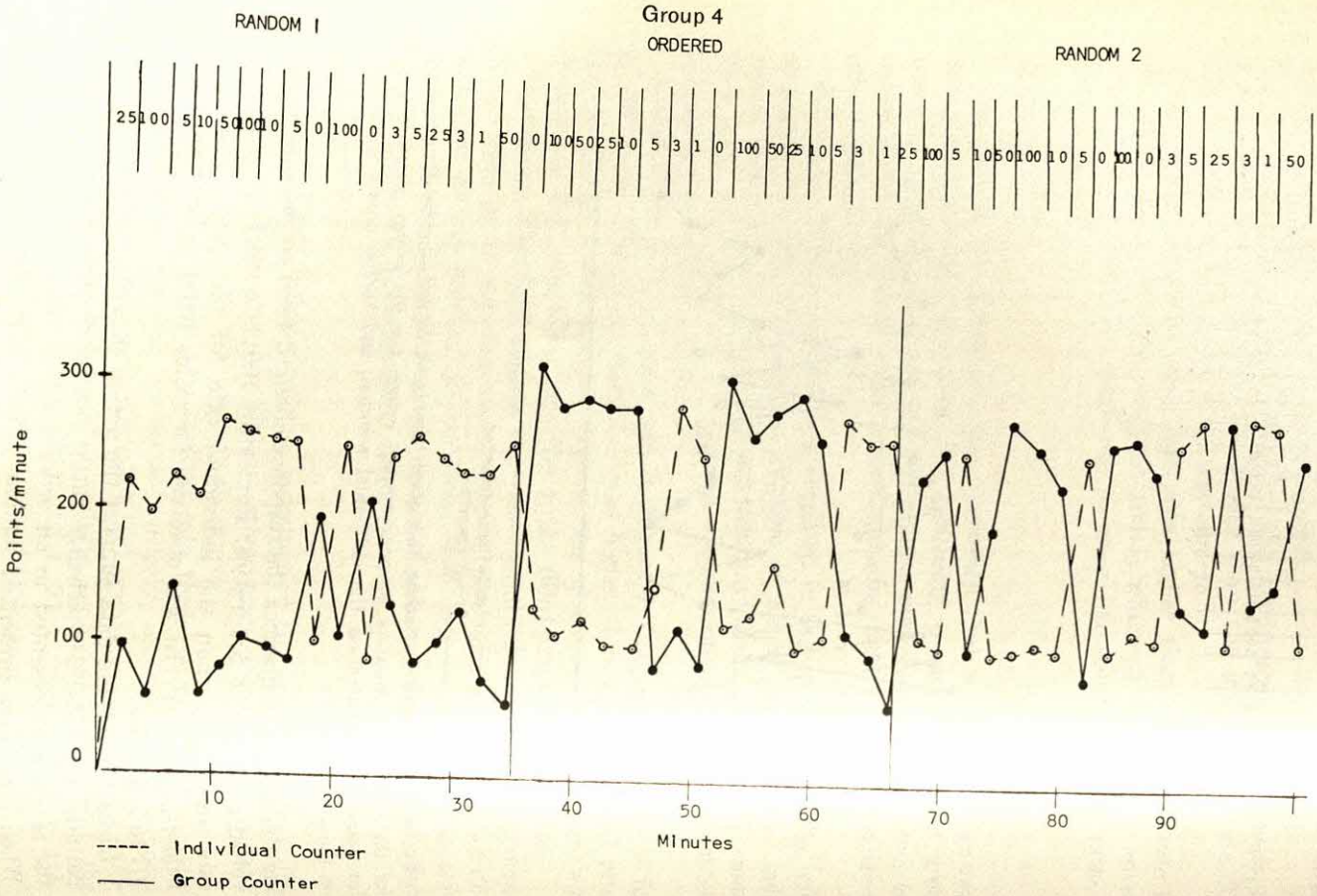


FIG. 10. Same as 9.

during the ordered presentations, it persisted in the subsequent random presentation.

SUMMARY AND CONCLUSIONS

The findings of this study are summarized in Table 2. They suggest that cooperation can be developed during group and individual reinforcement with individual feedback, if the rewards for achieving the group task are greater than for individual tasks. A preliminary investigation of different ratios for the two tasks suggests that discriminative

TABLE 2
Summary of Ratio Manipulations and Results for Experiments I-VII

Experiment	Group	Reward Ratios Investigated	Results
I	1	A. 10:1	No cooperation
		B. 1:1	No cooperation
		A. 10:1	No cooperation
II	1–4	A. 100:1	Cooperation
		B. 1:1	No cooperation
		A. 100:1	Cooperation
III	1	Ordered presentation of ratio sequences	
		I. 100:1, 50:1, 25:1, 10:1, 5:1, 3:1, 5:1, 5:1	Cooperation breaks down at 3:1
		II. 100:1, 50:1, 25:1. 10:1, 5:1, 3:1, 1:1, 3:1, 5:1, 10:1	Cooperation breaks down at 1:1
IV	1	Random presentation of ratios	Cooperation breaks down at 3:1
V	4	Random presentation of ratios	No cooperation
VI	4	Random presentation, 0¢ for individual counter to initiate cooperation	No cooperation
VII	4	A. Random	No cooperation
		B. Ordered - 0:1, 100:1, 50:1, 25:1, 10:1, 5:1, 3:1, 1:1	Cooperation breaks down at 3:1
		A. Random	Cooperation

response patterns for different ratios can be developed more readily when ratios are presented in some orderly sequence. This finding is in line with the more general hypothesis that the development of cooperation in alternative task situations is dependent upon whether or not the group is in a situation conducive to learning whether it is group or individual tasks that return the greater rewards.

REFERENCES

AZRIN, N. H., & LINDSLEY, O. R. The reinforcement of cooperation between children. *Journal of Abnormal Social Psychology*, 1956, **52**, 100–102.

This research investigates the function of similarity arising from matching colors. A modified matching-to-sample procedure allows similarity or dissimiliarity to be presented contingent on a response. A standard key is lit by one color on each trial. A response to one comparison key illuminates the key with the same color as the standard. A response to the other key produces a lamp illumination of dissimilar color. In this manner, similarity would be contingent on pressing one key and dissimilarity on pressing the other key. The purpose of this study was to determine whether young children would maximize similarity in a situation in which similarity was contingent on pressing one key and dissimilarity on pressing another key.

In a situation where *S* can produce only similarity or dissimilarity, a *S* who maximizes similarity will minimize dissimilarity. Hence, the effective stimulus could be either similarity or dissimilarity (or both). However, for simplicity of exposition this research is described in terms of similarity.

METHOD

Subjects

The *S*s, 15 boys and 15 girls, ranged in age from 50 to 68 months (mean = 61), and were attending the University of Iowa Preschools.

Apparatus

An 18-inch square response panel, mounted flush in a wall, contained three 2.0-inch diameter panels of Polacoat plexiglas. The panels formed an equilateral triangle with each panel separated from the others by 8.0 inches. The two panels which formed the base of the triangle were response keys which activated a switch when pressed, and the uppermost panel (standard panel) was immovable. Each panel could be transilluminated by either red, green, yellow, or white lamps. Programming and recording equipment was located in an adjoining room. A room air conditioner provided background noise.

Experimental Design

Similar stimulus condition. Twenty *S*s were run under a condition in which the standard was transilluminated either red or green and the response keys were dark. A response to one key produced a color identical to that of the standard, and a response to the other key produced a color different from the standard. Responses to the same key sometimes made the key red and sometimes made it green. Thus, one key was a matching key, and the other was a nonmatching key. Half the *S*s began

performance with the left key as the matching key and the remainder with the right key as the matching key. For all *Ss*, the matching key was switched half way through the experiment.

Different stimulus condition. Ten *Ss* were run under a condition which was identical to the similar stimulus condition except that a yellow color was substituted for the red standard color and a white for the green standard color. For half the *Ss*, the left key produced a red illumination if the standard was white and produced green if the standard was yellow. The right key produced red if the standard was yellow and green if the standard was white. The remaining *Ss* began with the right key producing red given a white standard and green given a yellow standard. For all *Ss* the key-stimulus combinations were reversed half way through the experiment.

This condition served as a control for "symbolic matching." Cumming & Berryman (1965) have used symbolic matching to refer to responding under the control of the standard stimulus when the standard is different from all comparison stimuli. In this experiment, if *S* were to respond to the left (right) key throughout the first half of the experiment and to the right (left) key the remainder then the pattern of responding would be symbolic matching because each standard color would be paired with only one comparison color.

The experimental design can be summarized as a mixed four factor design. The similar stimulus *vs* different stimulus groups, and the left key match *vs* right key match groups constituted the two between subject factors. The reversal of key color after Trial 64 for all *Ss* provided a training phase factor, and the trials within each phase were equally divided into three blocks for the trial blocks factor.

Procedure

Prior to entering the experimental room *S* selected and took possession of a toy which was offered for participation in the experiment. The *S* was seated in front of the apparatus on a cushion and was instructed to: (*a*) push one of the two bottom keys whenever the top panel was illuminated, and (*b*) play the game until the experimenter said, "stop." The experimenter sat behind *S* facing away from the stimulus panel. There were 128 trials for each *S*. A trial began with the illumination of the standard. The first subsequent depression of a response key produced the immediate illumination of that key. The trial ended 1.5 seconds later when the two illuminated panels were darkened. The intertrial interval between standard offset and onset of the new standard was 1.0 seconds. Responses after the first on each trial had no programmed consequence. In the event *S* kept the key depressed, the intertrial interval did not end

until the key was released. This contingency was introduced because it was assumed that key holding would be infrequent if it prevented illumination of the standard panel. An 8-bit recycling shift-register programmed the same sequence of standard colors for each *S*. The sequence provided an equal number of trials with each color and was arranged so that *S* would match colors on 50% of the trials if he switched keys after each response.

RESULTS

A comparison of performance under the similar stimulus condition and the different stimulus condition requires specification of the scoring procedure since only symbolic-matching responses were possible in the different stimulus condition. For each *S* in the dissimilar stimulus group, the number of responses to the left key for the first 64 trials plus the number of responses to the right key for the second 64 trials were divided by 128 (total trials). If the resulting proportion was equal to or greater than .5, then every response to the left (right) key in the first (second) training phase was scored as a symbolic-matching response. If the proportion was less than .5, then responses to the right (left) key in the first (second) training phase were scored as symbolic-matching responses. The total proportion of symbolic-matching responses for *S* could range from .5 to 1.0, but the proportion of matching responses for *S*s in the similar stimulus condition could range from .0 to 1.0. This scoring procedure, therefore, provided a conservative evaluation of whether the similar stimulus group emitted more matching responses than the dissimilar stimulus group. For the different stimulus condition the dependent variable of number of matching responses refers to the number of *symbolic*-matching responses.

Figure 1 presents mean matching responses for the similar stimulus and different stimulus groups for eight blocks of 16 trials. A mixed analysis of variance design composed of two between-subject factors (similar *vs* different stimuli, left key match *vs* right key match) and two within-subject factors (training phase, trial blocks) revealed a significant main effect of trial blocks, $F(3, 78) = 7.90$, $p < .001$, and a significant stimulus $\times$ trial blocks interaction, $F(3, 78) = 3.74$, $p < .025$. In followup analyses with the similar stimulus group, a significant trial blocks effect was obtained, $F(3, 54) = 9.06$, $p < .001$. A test for linear, quadratic, and cubic trends revealed only a significant linear comparison, $F(1, 57) = 50.29$, $p < .001$. Separate analyses contrasting the similar and dissimilar groups for each trial block produced a significant effect for only trial block three, $F(1, 26) = 5.88$, $p < .025$. Thus, the similar stimulus groups showed an increase in matching responses over trial blocks.

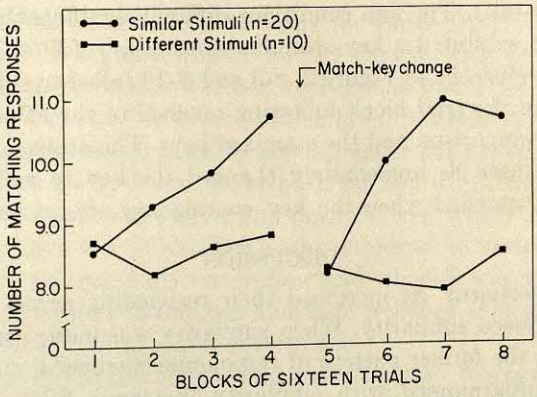


FIG. 1. Number of matching responses by trial blocks for the similar stimulus and different stimulus conditions.

In order to determine whether individual Ss maximized responding to the key programmed for similarity, a normal approximation to the binomial was applied to each S's proportion of matching responses (Hays, 1963). Figure 2 presents individual performances representative of the performances of the eight Ss in the similar stimulus group who maximized responding to the key programmed for similarity ($p < .01$, two-tail). No S in the different stimulus group performed beyond chance expectations

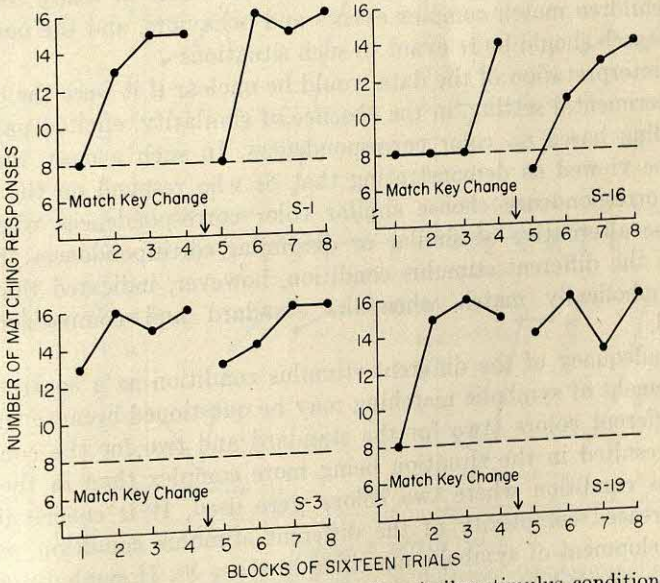


FIG. 2. Individual performance under the similar stimulus condition.

($p > .05$, two-tail). The two conditions differed significantly in the number of *S*s who exhibited a key-stimulus contingency ($X^2 = 3.60$, $p < .05$, one-tail). In reference to Figure 2, *S*-3 and *S*-19 failed to show a decrease in matching in the trial block following reversal of the color contingency between the comparison and the standard keys. The absence of a decrease was because these *S*s immediately changed the key to which they predominently responded when the key contingency was reversed.

DISCUSSION

In this experiment, *S*s increased their responding across trials to the key that produced similarity. When similarity was made contingent on a different key, the former pattern of responding weakened, and responding to the key programmed with similarity increased. Such a pattern of performance cannot be attributed to position responding because *S*s maintained matching performance when the key which produced similarity was changed. It can be concluded that similarity served as a positive reinforcer (and/or dissimilarity served as a punishing stimulus) because differential responding to the comparison keys was maintained by the consequence of similarity.

The fact that this experiment does not separate the effects of similarity and dissimilarity poses no particular problem. In any situation where *S* can produce only similarity or dissimilarity, the absence of similarity is the presence of dissimilarity. This condition exists in many settings in which children match complex events and behaviors, and the outcome of this research should be relevant to such situations.

The interpretation of the data would be unclear if it were the case that the experimental setting, in the absence of similarity, elicited patterns of responding based on color correspondences. In such a case, the results could be viewed as demonstrating that *S*s who respond on the basis of color correspondence choose similar color correspondences when faced with the alternative of similar or dissimilar correspondences. Performance in the different stimulus condition, however, indicated that *S*s did not symbolically match when the standard and comparison colors differed.

The adequacy of the different stimulus condition as a control for the development of symbolic matching may be questioned because the use of four different colors (two for the standard and two for the comparison keys) resulted in the situation being more complex than in the similar stimulus condition where two colors were used. It is conceivable that the increased complexity of the different stimulus condition prevented the development of symbolic matching in those *S*s. If symbolic matching could be elicited by simply instructing *S* to produce specific symbolic

matches, there would be no basis for considering the different stimulus condition to be too complex for the development of symbolic matching. In order to determine whether instructions would induce symbolic matching, ten four- and five-year-old children received explicit instructions to symbolically match prior to serving under the dissimilar stimulus condition. Each *S* was instructed to push the key which would be red when the standard panel was yellow and push the key which would be green when the standard was white. The *S* was requested to verbalize what he should do when the standard was yellow and when it was white. Five *S*s failed to verbalize the instructions and were replaced. The rest of the procedure was identical to the main experiment.

Figure 3 presents individual performances representative of the performances of the six *S*s who showed greater than chance level of symbolic matching ($p < .01$, two-tail). Of these six *S*s, four performed the symbolic match which had been instructed. Subject A2, Figure 3, was one of the two *S*s who produced the symbolic match which was not instructed. The mean frequency of symbolic match responses for these ten *S*s over trial blocks was parallel with, but slightly higher than, the similar stimulus group. The performance of these *S*s demonstrated that some four- and five-year-old children will symbolically match when so in-

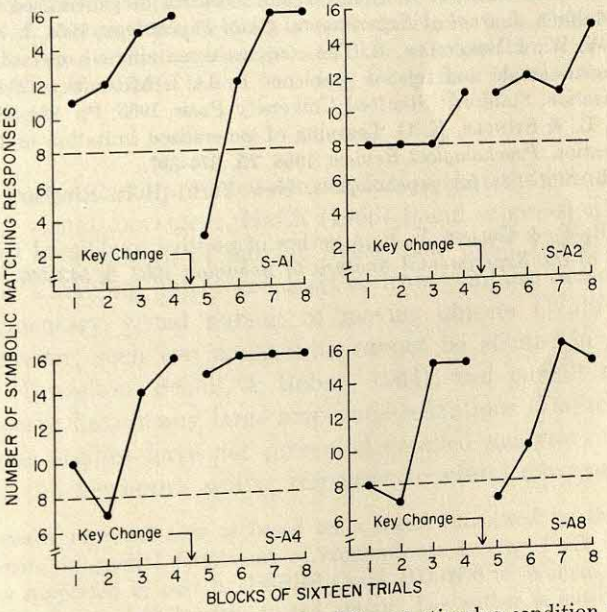


FIG. 3. Individual performance under the different stimulus condition for subjects instructed to symbolically match.

structed. Since symbolic matching can be elicited with simple instructions, it appears that task complexity does not account for the lack of symbolic matching in the different stimulus condition. Hence, the different stimulus condition was an appropriate control for determining the extent to which children develop symbolic matching in the absence of similarity.

This research found that similarity as a stimulus consequence was a reinforcing event. Since the training history which determined the function of similarity was not experimentally introduced, it is likely that the present experimental situation is only one of many situations in which similarity is a reinforcing stimulus for young children. Those situations may include both the matching of physical events and behaviors. Baer & Sherman (1964) have presented evidence which can be interpreted in terms of similarity reinforcing responses which match behaviors of a puppet. The present study indicated that with a modified match-to-sample task, the similarity of colors serves as a reinforcer. These two studies are consistent with the general thesis that for young children the reproduction of observed events is reinforced by the relational stimulus of similarity.

REFERENCES

Baer, D. M., & Sherman, J. A. Reinforcement control of generalized imitation in young children. *Journal of Experimental Child Psychology*, 1964, **1**, 37–49.

Cumming, W. W., & Berryman, R. The complex discriminated operant: Studies of matching-to-sample and related problems. In D. I. Mostofsky (Ed.), *Stimulus generalization*. Stanford: Stanford University Press, 1965. Pp. 284–330.

Gewirtz, J. L., & Stingle, K. G. Learning of generalized imitation as the basis for identification. *Psychological Review*, 1968, **75**, 374–397.

Hays, W. L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.

Kelleher, R. T., & Gollub, L. R. A review of positive conditioned reinforcement. *Journal of the Experimental Analysis of Behavior*, 1962, **5**, 543–597.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 469–482 (1969)

The Ocular Response of Human Newborns to Intermittent Visual Movement[1]

LYN W. WICKELGREN
Yale University

Two experiments were designed to assess the human newborn's ocular response to intermittent visual movement created by sequentially illuminated lights. In Experiment 1, photographs were made of 18 *S*s' eyes when they were presented a stationary light centered in front of the eyes, a blinking center light, and a repeating series of three horizontal lights (right, center, left, center, etc.). In Experiment 2, 18 *S*s were presented the same repeating series of horizontal lights at three different light durations. In both experiments, the two eyes generally did not converge upon one place; the left eye looked left of the center of the field, while the right eye looked right of center. Conjugate eye movements occurred about 45–50% of the time in all stimulus conditions. *Every S* in both experiments responded to the lights' movement by shifting direction of gaze toward the left light when it went on and toward the right light when it went on. The joint occurrence of nonconvergence and frequent conjugate eye movements may explain the absence of a strong shift tendency toward the center light. Only 5 *S*s followed the lights across the stimulus field. Rate of intermittent visual movement within the limits tested here was not critical for newborn ocular responding.

Several studies have provided some evidence that the human newborn responds to visual movement. Haith (1966) found suppression of sucking in response to a pattern of moving lights, and Wolff (1959; 1965) observed that a moving object can alert a drowsy infant. Some newborns show rudimentary visual pursuit of moving objects (Wolff & White, 1965). However, such eye movements cannot be elicited in all normal newborns (Brazelton, Scholl, & Robey, 1966), and pursuit movements which do occur have many large amplitude refixations (Dayton & Jones, 1964). These studies have not presented detailed analysis of the components of the newborn's ocular responses to visual movement largely

[1] The research reported here is based on a thesis submitted by the author to Yale University in partial fulfillment of requirements for the Ph.D. degree. The research was supported in part by USPHS Grant HD-0890 to William Kessen, and by an NSF Predoctoral Fellowship to the author. The author is indebted to William Kessen, Marshall M. Haith, Mrs. Peggy Yarsawich, and the nursing staff of the maternity ward of the Yale-New Haven Hospital.

because of the limitations of electro-oculography and of reports by human observers.

The present study consisted of two experiments designed to assess in detail the response of newborn infants to visual movement. The infant's ocular response to sequentially blinking lights was recorded photographically. The first experiment asked how, if at all, the newborn responded ocularly to visual movement represented by change in stimulus position. The second experiment tested the effect of rate of visual movement, specifically, whether or not the precision of the ocular response to intermittent visual movement would increase with increased light duration. The experiments also assessed both the extent to which each *S*'s two eyes were directed toward the same place at a given moment (binocular convergence) and the frequency with which each *S*'s two eyes moved simultaneously in the same direction (conjugate eye movements).

METHOD

Subjects

Subjects for the experiments were 36 awake full-term newborn infants from 19.5 to 134 hours of age in the nursery of the Yale-New Haven Hospital. In Experiment 1 were 13 males and 5 females; in Experiment 2 were 10 males and 8 females. An additional 65 babies were seen, but they could not be used as *S*s either because they fell asleep during the observation or because their film records were unscorable.

Apparatus

The experiments were conducted in a small light-attenuated room on the maternity ward of the hospital. Figure 1 shows a diagram of the apparatus. The *S* was placed face upward in a head-restraining crib on a table. The stimulus panel was mounted 9½ inches above the infant's head. The three stimulus lights were Dialco #26410–115 fitted with GE 1819 bulbs. They were horizontally placed in the panel (left, center, and right). The lights were covered with tissue paper so that *S* saw white lighted circles $^{11}/_{16}$ inch in diameter. The lights were centered 3 inches apart, and the total stimulus field, 6 inches across, subtended approximately 34° 28′ of visual angle. The brightness of each stimulus light was approximately 7 foot-lamberts, measured from the position of *S*'s eye by an SEI Exposure Photometer.

An Automax Model G-2 35-mm variable-speed soundproofed camera was mounted above the stimulus panel holder (see Figure 1). The camera was fitted with a 105-mm Takumar lens and an SII adaptor ring. Kodak high-speed infrared film (HIR 417) was used for photographing the eyes.

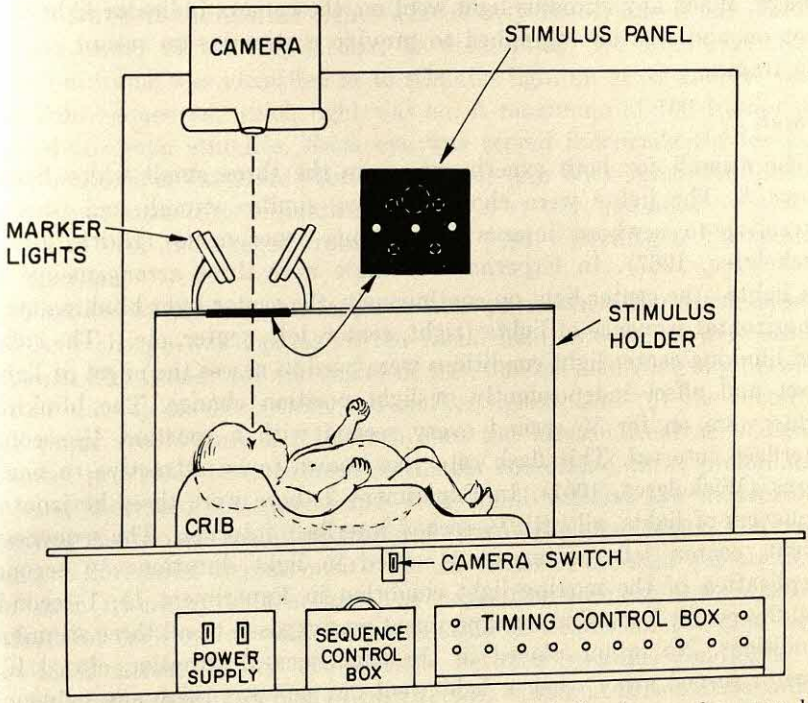


FIG. 1. Side-view diagram of apparatus with inset of stimulus panel as seen by *Ss*. Panel contains three white stimulus lights and holes behind which are marker lights and camera.

Photographs were made at *f*/8.0 with a camera focal distance of 27 inches. The field of the photographs encompassed *S*'s head, which was centered directly below the lens of the camera. Photographs were made twice per second with an exposure of 1/6 second. A counter recorded the number of film frames exposed.

Four Bausch and Lomb Nicholas illuminators were fitted with Kodak Wratten 87C and Corning C.S. 7–69 filters; reflections of these infrared lights on the eyes were visible on film and were used as marker lights to assess the orientation of the eyes. The filters eliminated all visible light except a barely visible reddish beam above 800 mμ.

The control apparatus for the stimulus lights (see Figure 1) was supplied by a 28-volt D.C. regulated power supply. A sequence-control rotary switch was wired to make available different sequences of lights. A relay system provided timing control so that both the duration of light flashes and the interflash interval could be varied. Since the system controlling the light flashes was independent of the system controlling the camera shutter, the stimulus lights were connected to an indicator light in the

camera. When any stimulus light went on, the camera indicator light also went on and was photographed to provide a stimulus-on record on the film frame.

Stimuli

The stimuli for both experiments were the three small white lights above *S*. The lights were chosen because similar stimuli had proved attractive to newborn infants in previous experiments (Haith, 1966; Wickelgren, 1967). In Experiment 1 there were three arrangements of the lights—the center light on continuously, the center light blinking, and a horizontal sequence of lights (right, center, left, center, etc.). The solid and blinking center light conditions were used to assess the effect of light onset and offset independently of light position change. The blinking lights were on for ½ second every second with a constant ½-second interflash interval. This flash rate was known to be attractive to newborns (Wickelgren, 1967). In Experiment 2 there were three horizontal sequences of lights, all with ½-second interflash intervals. The sequences (right, center, left, center, etc.) varied in light duration: ½ second (replication of the moving light condition in Experiment 1), 1 second, and 2 seconds. Each *S* in an experiment was exposed to all three stimulus conditions. No infant placed in the experimental situation closed his eyes or turned away when a light went on; nor was there any evidence of startle when a light went on.

Procedure

The same general procedure was followed in both experiments for all *S*s. Subjects were observed about 30 minutes before the morning feeding, between 8:30 and 9:30 AM. The experimenter brought *S* to the observation room from the nursery and placed him in the crib. The *S*'s head was centered under the camera, and a blind nipple was placed in his mouth. The center light was turned on, and the room lights were turned off. The camera was turned on, and *S* received successive presentations of the three stimuli according to a predetermined order counterbalanced across *S*s. Moving light sequences always began with the right light. Each stimulus condition was presented at least 50 seconds to allow 100 photographs of the eyes. The center light was on for 2 to 3 seconds between stimulus presentations to prevent a totally dark stimulus field. If *S* closed his eyes, *E* tried to rouse him but usually had to terminate the session.

Scoring and Analysis

No *S*'s film was scored unless a preliminary check indicated that there was a minimum of 75 scorable frames for each stimulus. A film frame

was scorable if the marker lights were clearly visible and if the position of the center of the pupil could be determined (see Figure 2). Each scorable frame was identified as to stimulus light on or off and, for moving light sequences, which light was on. A maximum of 100 frames was scored for each stimulus. Each eye was scored independently for each film frame on a Vanguard Motion Analyzer (see Wickelgren, 1969). Both inter- and intrascorer reliabilities (Pearson r) were obtained for several strips of film and ranged from +.92 to +.96, according to the quality of the film.

Location of ocular orientation: convergence. In order to determine where each eye was looking in the visual field, two points were scored on each film frame: (*a*) the center of the stimulus field (halfway between either the top or the bottom two marker lights), and (*b*) the center of the pupil (see Figure 2). In each case the Motion Analyzer's digital counter provided the horizontal rectilinear coordinate. Since preliminary scoring indicated infrequent and very small vertical eye movements, vertical coordinates were not obtained (see Wickelgren, 1969). The pupil center's horizontal displacement from the center of the field was obtained for each frame by subtraction. The mean displacement score was obtained for each eye of each *S* in each stimulus condition. The data were combined across *S*s for comparisons between stimulus conditions as preliminary analysis indicated no effect of stimulus presentation order. The resulting mean displacement scores indicated the average amount of convergence for each stimulus condition.

Conjugate eye movements. Films were next scored to determine the frequency of conjugate eye movements. For each film frame on which

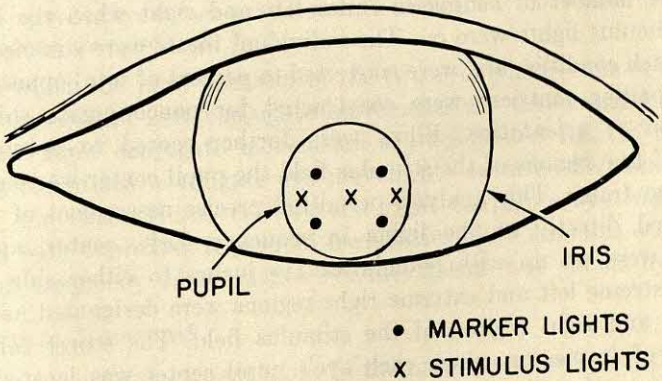


FIG. 2. Diagram of *S*'s eye. The four dots represent filmed reflections of the marker lights on the cornea; the three x's indicate positions of the stimulus lights relative to the marker lights. Stimulus lights are not visible on film.

both eyes were scorable, the joint behavior of the two eyes between successive frames was noted. There were nine possible joint events which might occur between successive film frames because either eye could move left, move right, or remain stationary. Only two of the nine categories, both eyes moving left and both eyes moving right, represented conjugate eye movements. Tallies were made for each *S* in each stimulus condition of the number of times the joint behavior of the two eyes was in each category. The nine totals for all *S*s in each stimulus condition were converted to percent of the overall total for that condition.

Subjects' ocular responses to visual movement were assessed at three levels of analysis:

Variability of ocular orientation. The standard deviation of the displacement scores (see above) was obtained for each eye of each *S* in each stimulus condition. The data were combined across *S*s for comparisons between stimulus conditions. The resulting standard deviations of the displacement scores provided a general measure of whether or not overall dispersion of ocular orientation was related to intermittent visual movement.

Shifts in ocular orientation. The number and direction of shifts in location of ocular orientation (i.e., displacement scores) between successive frames were scored. First, the number of left and right shifts between frames was tallied for each eye of each *S* in each stimulus condition. Since the number of scorable frames varied among *S*s, the totals obtained were converted to proportions; means were obtained. Next, the direction of shifts between film frames was determined for moving light conditions to assess whether *S*s were responding to the intermittent movement of the lights by shifting direction of gaze toward them. A matrix was constructed for each *S* in each moving light condition. The entries were the number of conjugate shifts left and right when the left and right stimulus lights were on. The individual totals were summed across *S*s for each condition and were converted to percent of total opportunities. Corresponding matrices were constructed for nonconjugate shifts.

Region of orientation. Films were further scored to determine in which of five regions of the stimulus field the pupil center was located on each film frame. This analysis permitted precise assessment of whether *S*s looked directly at the lights in sequence. Left, center, and right regions were set up with boundaries 1½ inches to either side of each light. Extreme left and extreme right regions were designated as beyond the left and right regions of the stimulus field. The scorer tallied the number of frames on which each eye's pupil center was located within each region when the different lights were on (see Wickelgren, 1969, for details about calibration).

EXPERIMENT 1

Results

Location of Ocular Orientation

The two eyes never looked at the same place on the stimulus field in any stimulus condition; the left eye consistently looked left of center, while the right eye looked to the right of center. This lack of convergence between the two eyes was highly significant ($F = 63.36$, $df = 1,34$, $p < .001$). In all three stimulus conditions, the two eyes were oriented an average of about 4½ inches apart at the stimulus plane. Convergence upon a single light would require that the eyes be oriented no more than $^{11}/_{16}$ inch apart.

Conjugate Eye Movements

The frequency of conjugate eye movements was assessed for all *Ss* combined in each stimulus condition. Table 1 shows the percent of frames

TABLE 1
PERCENT FRAMES IN EACH CATEGORY OF JOINT EYE BEHAVIORS FOR EACH STIMULUS CONDITION, EXP. 1

Stimulus condition:	Categories[a] —	–L	–R	L–	R–	LR	RL	LL	RR	Total %
M	7.6	8.3	9.3	8.0	8.3	5.6	6.1	22.6	24.2	= 100.0
B	9.2	8.5	9.8	7.9	9.4	6.9	6.9	20.6	20.8	= 100.0
S	8.0	6.3	8.8	7.7	9.3	6.7	7.0	23.8	22.4	= 100.0

[a] In each pair of symbols, the left symbol indicates the behavior of the left eye between frames, and the right symbol indicates the behavior of the right eye between the same frames. L represents eye movement to the left; R represents eye movement to the right; – represents no eye movement.

between which the joint behavior of the two eyes was in each of the nine possible categories. Left conjugate movements are designated LL in the table, and right conjugate movements are designated RR. In all three stimulus conditions, conjugate eye movements occurred between 40 and 45% of the time. Moreover, left and right conjugate eye movements occurred with approximately equal frequencies in all stimulus conditions. The high frequency of conjugate eye movements was shown by all *Ss*.

Responses to Movement

Variability of ocular orientation. In the movement condition, the mean standard deviation of ocular orientation was .030 for the right eye and .029 for the left eye. In the blink condition, the mean standard deviation

was .025 for each eye; in the solid condition, .026 for each eye. There was significant differential variability in location of ocular orientation across stimulus conditions ($F = 5.00$, $df = 2,68$, $p < .01$). The variability was significantly greater in the movement condition than in both the blink condition ($p < .01$) and the solid condition ($p < .05$) (Newman-Keuls test). There was not a significant difference between the blink and solid conditions. Thus, light onset-offset alone does not account for the high dispersion of location of ocular orientation in the movement condition; position change of the lights was responded to independently of light onset and offset. There was no difference between the two eyes in variability of orientation and no interaction.

Shifts in ocular orientation. Both eyes generally shifted position between about 75% of the frames in all stimulus conditions. Furthermore, there was no differential tendency for either eye to shift more in one direction than in the other; that is, both eyes shifted left and right equally often in the three conditions.

Every *S* responded appropriately to the moving lights by shifting direction of gaze predominantly toward the left and right lights when they were on. The shift directions were statistically analyzed for both conjugate and nonconjugate eye movements. When the left light was on, conjugate shifts were toward the left 80% of the time, and when the right light was on, conjugate shifts were toward the right 79% of the time. A dependent t test substantiated this finding; there were significantly more left conjugate shifts when the left light was on than when the right light was on (mean difference = .60, $t = 8.44$, $df = 17$, $p < .001$). Because the proportions of right conjugate shifts are the arithmetic inverse of the left proportions, the t test finding is equivalent to asserting that there were also more right conjugate shifts when the right light was on than when the left light was on.

When the left light was on, nonconjugate left shifts occurred 56% of the time, and when the right light was on, nonconjugate right shifts occurred 59% of the time. There were significantly more left nonconjugate shifts when the left light was on than when the right light was on ($F = 4.19$, $df = 1,34$, $p < .05$). There was no difference between the two eyes and no interaction.

It should be emphasized that the occurrence of appropriate directional shifts does not imply that the eyes shifted to the left or right of the center of the field or that the eyes looked directly at the appropriate light after shifting toward it. The shifts were made to the left or right of the eyes' previous position without regard to where that position was in the stimulus field.

There was high individual variability in number and direction of shifts during the interval after each stimulus light went off. When the center light was on, *Ss* did not reliably shift direction of gaze toward it.

Region of orientation. Most of the *Ss*' displacement scores rested within the left, center, and right regions. Furthermore, most *Ss*' eyes shifted predominantly within one region, even in the movement condition. There was no evidence of improvement in following the lights between regions during exposure to the moving lights. Two of the 18 *Ss* did follow the lights between regions in the movement condition. During the right-center-left-center sequences, their eyes shifted to the right, center, left, and center regions in turn. Figure 3 compares typical regional analysis records of a follower and a nonfollower. Both followers shifted direction of gaze consistently throughout most of the stimulus presentation. Moreover, they both showed an above-average frequency of conjugate eye movements in following the lights. Whereas the other *Ss* had conjugate eye movements about 45% of the time in the movement condition, the followers showed conjugate eye movements 60% of the time. Age, sex, and birth conditions did not appear to differentiate the followers from the nonfollowers.

EXPERIMENT 2

The results of Experiment 1 indicated that all *Ss* were able to respond consistently to the moving lights by shifting direction of gaze toward both the left and right lights. Although the *Ss*' eyes moved toward the lights, they generally did not shift far enough to look directly at the lights and follow the lights in sequence. Since the light duration in Experiment 1 was ½ second, it was possible that most *Ss* did not have time to respond fully to the moving lights and that a longer light duration might prove more effective in eliciting a precise response. Thus, two longer durations were selected for sequences in Experiment 2, 1 second and 2 seconds (still longer light durations lose the appearance of a smooth sequence). A sequence with ½-second flash duration was used to replicate the movement condition findings of Experiment 1.

Results

Location of Ocular Orientation

As in Experiment 1, the two eyes did not converge, i.e., look at the same place on the stimulus field in any stimulus condition. The left eye consistently looked to the left of center, while the right eye looked right

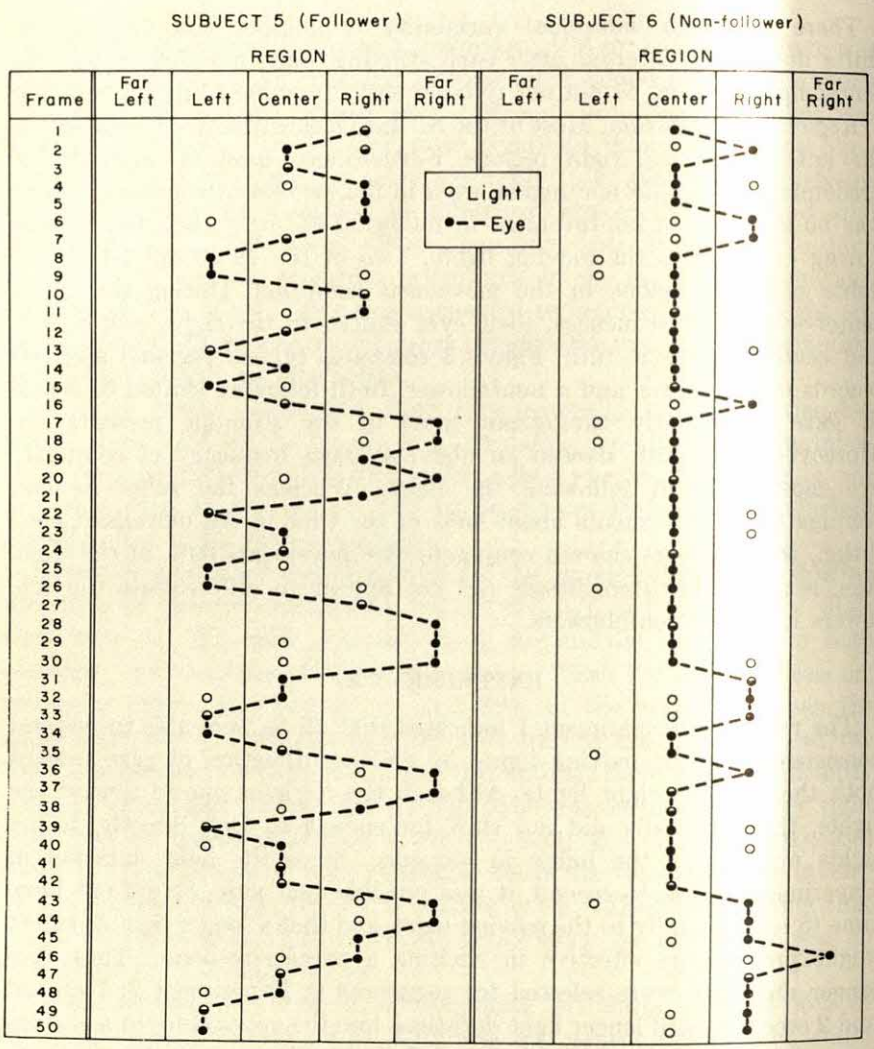


FIG. 3. Two records of region of ocular orientation in Experiment 1. Left is a record of one *S* who followed the lights between regions of the field; right is a record of a typical *S* who did not follow the lights between regions. The records cover the first 50 frames for the right eye in the movement condition. A circle outline indicates which light—left, center or right—was on during each frame; no circle means no light was on during the frame. Lines connect solid black circles to indicate the sequence of the *S*s' eye movements from frame to frame.

of center. This difference between the two eyes was highly significant ($F = 97.33$, $df = 1,34$, $p < .001$). In all stimulus conditions, the eyes were oriented an average of 5 inches apart at the stimulus plane, and the eyes were centered about the center of the field.

Conjugate Eye Movements

Table 2 shows the percent of frames between which the joint behavior of the two eyes was in each of the nine possible categories. In all three stimulus conditions, conjugate eye movements occurred about 50% of the time. Moreover, left (LL) and right (RR) conjugate eye movements occurred with approximately equal frequencies in all stimulus conditions, as in Experiment 1. The high frequency of conjugate eye movements was shown by all *S*s.

Responses to Movement

Variability of ocular orientation. The mean standard deviations of location of ocular orientation in both the ½-second and 1-second conditions were .032 for the right eye and .034 for the left eye. Corresponding figures for the 2-second condition were .037 for both eyes. There was

TABLE 2

PERCENT FRAMES IN EACH CATEGORY OF JOINT EYE BEHAVIORS FOR EACH STIMULUS CONDITION, EXP. 2

Stimulus condition:	Categories[a] —	–L	–R	L–	R–	LR	RL	LL	RR	Total %
½	7.2	7.2	7.2	8.6	6.6	4.9	6.8	25.6	25.9	= 100.0
1	7.6	7.8	7.2	9.1	7.6	5.7	6.1	25.2	23.6	= 99.9
2	7.3	7.9	7.3	9.3	7.4	5.2	5.6	25.0	24.9	= 99.9

[a] In each pair of symbols, the left symbol indicates the behavior of the left eye between frames, and the right symbol indicates the behavior of the right eye between the same frames. L represents eye movement to the left; R represents eye movement to the right; – represents no eye movement.

significant differential variability across stimulus conditions ($F = 6.67$, $df = 2,68$, $p < .01$). The variability was significantly greater in the 2-second condition than in either the ½-second ($p < .01$) or the 1-second condition ($p < .01$) (Newman-Keuls test). There was no difference between the two eyes and no interaction.

Shifts in ocular orientation. The eyes generally shifted position between about 78% of the frames. As in Experiment 1, there was no difference between the two eyes or the stimulus conditions in mean proportion of shifts, and there was no differential tendency for either eye to shift more in one direction than in the other in any condition.

As in Experiment 1, every *S* responded appropriately to the moving lights by shifting direction of gaze predominantly toward the left when the left light was on and toward the right when the right light was on. When the left light was on in any stimulus condition, conjugate shifts

were toward the left 71% of the time, and when the right light was on, conjugate shifts were toward the right 75% of the time. This direction-of-shift effect was highly significant ($F = 94.01$, $df = 1,17$, $p < .001$). There was no stimulus effect or interaction.

When the left light was on, nonconjugate shifts were toward the left 57% of the time, and when the right light was on, nonconjugate shifts were toward the right 52% of the time. There were significantly more left nonconjugate shifts when the left light was on than when the right light was on ($F = 8.77$, $df = 1,34$, $p < .01$). There was no other significant effect. Thus, in all three stimulus conditions when the left light was on, both conjugate and nonconjugate shifts were predominantly toward the left, while when the right light was on, shifts were predominantly toward the right.

In the interval after a stimulus light went off, there was high individual variability in shifts. When the center light was on, as in Experiment 1, *Ss* did not reliably shift toward it.

Region of orientation. Subjects showed the same general pattern of responses found in the first experiment. Most *Ss'* eyes did not consistently follow the lights between regions and look directly at the lights in any stimulus condition, nor was there any evidence of improved ocular following between regions during exposure to any moving light sequence. Three of the 18 *Ss* did follow the lights between regions. Instead of showing improved following with increased duration of light flash, however, these 3 *Ss* followed the lights regardless of light duration. Moreover, like the 2 followers in Experiment 1, the 3 *Ss* showed an above-average frequency of conjugate eye movements in all three stimulus conditions. Whereas the other *Ss* had conjugate eye movements about 50% of the time, the followers showed conjugate eye movements 60–70% of the time. No other characteristics appeared to differentiate the followers from the nonfollowers. The follower's record in Figure 3 is typical of their performance.

DISCUSSION

All infants in both experiments responded ocularly to the intermittent visual movement of the lights. Most dramatically, *all Ss* responded to the moving lights by shifting direction of gaze appropriately toward the left and right lights. It is unusual to find such uniformity among newborn infants in their responding. This striking consistency in responding may be an indication of the power of moving as opposed to stationary stimuli for newborn infants.

Given that *Ss* did respond to the left and right lights, it might appear

puzzling that a similar shift effect did not occur with the center light. However, since the gaze of each *S*'s two eyes typically straddled the center light, the eyes received competitive directional information from it. If one eye were dominant, that eye should have consistently responded to the center light by shifting toward it. There was no indication that such was the case. If, as seems more likely, both eyes were equally responsive to the visual stimulation, they would have had to shift in opposite directions to shift toward the center light simultaneously. Such cases of the left eye shifting right while the right eye shifted left (RL on the tables) were rare, while conjugate movements were frequent. Therefore, the combination of nonconvergence and conjugate eye movements undoubtedly prevented appropriate shift responses to the center light. If one eye were occluded in this experimental situation, one might predict that the open eye would shift toward all three lights in sequence. Since normally the newborn infant has both eyes open, and since his eyes usually do not converge, peripheral stimuli, which send a similar directional message to both eyes, well may be stronger for him than are central stimuli. The results of these experiments are certainly compatible with this view.

In Experiment 1, although the infants shifted direction of gaze toward the moving lights, the shifts of most *Ss* were not large enough to result in their looking directly at the lights. It seemed reasonable that *Ss* may not have followed the lights across the stimulus field because they needed more time to respond. Therefore, in Experiment 2 the duration of light flashes was increased. However, the increased light durations did not increase the precision in shift responses. Another possibility, which was not tested, is that the on-off quality of the lights disrupted responding. If, therefore, there were no interflash intervals between successive lights or if a single light were moved across the field, perhaps smooth sequential responding would occur. It is significant, however, that 5 *Ss* did consistently follow the lights between regions of the stimulus field. Further, the 3 *Ss* who followed the lights in the second experiment did so at all three durations. The 5 *Ss*' ability to follow the lights soon after birth and their above-average frequencies of conjugate eye movements suggest that other babies will develop a similar level of responding within the first weeks of life.

The direct, photographic analysis of conjugate eye movements in these experiments confirms and extends the findings of Dayton, Jones, Steele, and Rose (1964), who measured newborns' eye movements indirectly with electro-oculography. That study demonstrated, in contrast to common belief, that newborn infants can move their eyes conjugately. In the present study, *all Ss* were able to move their eyes conjugately.

and they did so about half of the time—quite a high frequency when one considers the large number of alternatives to conjugate eye movements.

Coexistent with frequent conjugate eye movements the newborns in this study showed marked divergence of ocular orientation. In a previous experiment (Wickelgren, 1967), human newborns' eyes were shown to diverge when a stationary stimulus (a square of color or a panel of stripes) was presented. The extent of divergence was not related to the visual stimulation. The present study extends those results to moving light stimuli, strengthening the conclusion that at birth human beings physically cannot consistently converge the two eyes. Since newborn infants show nonconvergence concurrent with frequent conjugate eye movements, these two responses probably reflect two independently developing physiological systems.

Hebb (1949) has postulated that eye movements are critical for the development of perception. This study did show that eye movements are well coordinated at birth with a high frequency of conjugate movements. Further, the study showed that such eye movements can be directed toward a moving stimulus in the visual field. At the same time, however, the eyes do not usually converge upon the stimulus. Thus, perhaps eye movements themselves may provide a useful index of early perceptual development.

REFERENCES

Brazelton, T. B., Scholl, M. L., & Robey, J. S. Visual responses in the newborn. *Pediatrics,* 1966, **37,** 284–290.

Dayton, G. O., & Jones, M. Analysis of characteristics of fixation reflex in infants by use of direct current electrooculography. *Neurology,* 1964, **14,** 1152–1156.

Dayton, G. O., Jones, M., Steele, B., & Rose, M. Developmental study of coordinated eye movements in the human infant. II. An electro-oculographic study of the fixation reflex in the newborn. *Archives of Opthalmology,* 1964, **71,** 871–875.

Haith, M. M. The response of the human newborn to visual movement. *Journal of Experimental Child Psychology,* 1966, **3,** 235–243.

Hebb, D. O. *The organization of behavior. A neuropsychological theory.* New York: Wiley, 1949.

Wickelgren, L. W. Convergence in the human newborn. *Journal of Experimental Child Psychology,* 1967, **5,** 74–85.

Wickelgren, L. W. *The ocular response of human newborns to intermittent visual movement* (Doctoral dissertation, Yale University) Ann Arbor, Mich.: University Microfilms, 1969. No. 69–8455.

Wolff, P. H. The development of attention in young infants. *Annals of the New York Academy of Sciences,* 1965, **118,** 815–830.

Wolff, P. H. Observations on newborn infants. *Psychosomatic Medicine,* 1959, **21,** 110–118.

Wolff, P. H., & White, B. L. Visual pursuit and attention in young infants. *Journal of the American Academy of Child Psychiatry,* 1965, **4,** 473–484.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 483–493 (1969)

Factors Affecting Lateral Differentiation in the Human Newborn[1]

GERALD TURKEWITZ, TINA MOREAU, LINDA DAVIS, AND HERBERT G. BIRCH
Albert Einstein College of Medicine

The effect of asymmetry of muscle tonus and of prior somesthetic stimulation on responsiveness to subsequent somesthetic perioral stimulation was examined. Infants were subjected to; asymmetry of tonus ($n = 14$), somesthetic stimulation of the right cheek ($n = 14$) or the left cheek ($n = 15$), or a control condition in which both types of asymmetry were eliminated ($n = 13$). All experimental treatments significantly increased responsiveness to both right and left stimulation. Since asymmetry of tonus and prestimulation of the right cheek produced significantly more responses to right than to left stimulation, it is likely that both factors contribute to the normally occurring lateral differences in responsiveness. Somesthetic priming of the two sides did not have opposite effects, indicating additional determinants of lateral differentiation.

The results of a number of recent investigations have indicated that the normal human newborn is asymmetrical in his responsiveness to stimulation. He responds more readily when somesthetic, auditory, or visual stimuli are presented at his right than when they are presented at his left (Siqueland, 1964; Turkewitz, Gordon, & Birch, 1965; Turkewitz, Birch, Moreau, Levy, & Cornwell, 1966; Turkewitz, Moreau, & Birch, 1968; Wickelgren, 1967). Since such early lateral differences may contribute to the subsequent development of lateral dominance (Gesell & Ames, 1947), lateral preference and hemispheric differentiation, an understanding of the factors affecting lateral differentiation of behavior in the infant, is potentially important for understanding the development of lateralization.

Previous reports from our laboratory have shown that lateral differences in the infant's responsiveness to laterally presented auditory and somesthetic stimuli are associated with his asymmetrical head position prior to stimulation, with lateral differences in responsiveness

[1] This investigation was supported in part by the National Institutes of Health, National Institute of Child Health and Human Development (HD-00719); by the Association for the Aid of Crippled Children; and by the National Association for Retarded Children. We thank Miss Susan Fleischer for her assistance in collecting the data, and Mrs. Sarah Winters and the nurses of the Well-Baby Unit of the Bronx Municipal Hospital Center for their cooperation.

eliminated by maintaining his head in a midline position for 15 minutes (Turkewitz, Moreau, & Birch, 1966; Turkewitz, Moreau, Birch, & Crystal, 1967). In one of these studies (Turkewitz *et al.*, 1966), we suggested that since the infant's characteristic posture with his head turned to the right results in the partial occlusion of his right ear, adaptation to a lower level of ambient auditory input at this ear could have produced a reduction in its auditory threshold. A parallel explanation was offered for the effect of the infant's prior asymmetrical head position on lateral differences in responsiveness to somesthetic stimulation (Turkewitz *et al.*, 1967), with the infant's characteristic head position resulting in more rather than in less somesthetic stimulation of the right side. Consequently, in this case, it was argued that the particular mechanism underlying lateral differentiation could be the priming effect of prior stimulation rather than the adaptational effect of prior insulation.

However, at the time these explanations were advanced, it was pointed out that an alternative explanation is possible. In addition to its contribution to lateral differences in the ambient levels of somesthetic and auditory input, the characteristic asymmetrical head position of the infant could also result in lateral differences in muscle tonus. This difference might then result in an increased motor readiness to turn to the right even in the absence of applied stimulation or it could, by virtue of its afferent consequences, contribute summatively to subsequent input at this side. Since keeping the head in the midline not only reduces asymmetry of stimulation but also reduces asymmetry of muscle tone, it was not possible in the earlier studies to determine whether the experimentally induced changes in responsiveness were based upon the modification of prior asymmetries of ambient stimulation at the two sides upon changes in the asymmetry of muscle tone, or upon the interaction of both factors. The present investigation was designed to permit a separate consideration of the effects of asymmetry of tonus and of prestimulation in order to determine whether the infant's asymmetrical head position results in lateral differences in responsiveness because it produces lateral differences in somesthetic input, or because it results in lateral differences in muscle tonus.

METHOD

Head turning to a laterally presented somesthetic stimulus was used as the indicator of lateral differences in responsiveness. To determine the separate effects of prior asymmetry of input and asymmetry of muscle tone on lateral differences in responsiveness, the following four groups of infants were studied; an altered tonus group, right and left somesthetic prestimulation groups, and a control group. Infants in the tonus

group were subjected to a condition designed to produce asymmetry of muscle tonus only; infants in the somesthetic groups were subjected to conditions designed to produce asymmetry of somesthetic input without associated asymmetry in muscle tonus; infants in the control group were treated in a manner designed to eliminate asymmetries of both somesthetic input and muscle tone. Following these different conditions, all infants were given the same test of responsiveness to somesthetic stimulation.

Subjects. The 56 infants studied were randomly assigned to the four groups. The sample was drawn from the population of full-term newborns in the well-baby nurseries of the Bronx Municipal Hospital Center. Only infants whose medical histories indicated no pre- or perinatal complications and whose Apgar scores at 1 minute after birth (Apgar & James, 1962) were 8 or higher were studied. The birth weights of the infants ranged from 2.40 to 4.04 kg (mean = 3.20). At the time of testing, the infants ranged in age from 25 to 61 hours (mean = 41.5). Because of possible changes in irritability following circumcision, only uncircumcised male infants were included, resulting in a sample having more female than male infants. There were 10 females and 3 males in the control group, 11 females and 3 males in the tonus group, 13 females and 1 male in the somesthetic right group, and 12 females and 3 males in the somesthetic left group.

EXPERIMENTAL PROCEDURE

Because responsiveness may vary with time of day and feeding condition, all babies were studied 45–75 minutes before the regularly scheduled 1:30 PM feeding. At the start of the pretest period, each infant was placed in his own bassinet in a relatively quiet laboratory. In order to prevent the infant from differentially stimulating the two sides of his face by his own arm movements, all infants were swaddled. To prevent differential stimulation from the substrate, the mattress was shifted and the infant's head was held so that it extended beyond the edge of the mattress.

The pretest period for each group lasted 15 minutes. During this period, infants in the tonus group had their heads held so that the right cheek was on the same level as the mattress but not in contact with it. Infants in both somesthetic groups had their heads held in a midline position and were intermittently stimulated by the *E*, who moved the ball of his thumb in an elliptical pattern across the infant's cheek from the corner of his mouth to his ear. Each pretest stimulus lasted for approximately 1, 2, or 3 seconds with the infant receiving a total of 20 seconds of stimulation during each minute of the pretest period. Within these limits, the sequence

of stimulation used was random. The times at which the pretest stimuli were to be presented and terminated were indicated to the *E* by a 15-minute tape recording. The same sequence of pretest stimulation was used for each infant in the two somesthetic groups. The only difference between these two groups was that infants in the somesthetic right group were prestimulated only on the right cheek and those in the somesthetic left group were prestimulated only on the left cheek. Infants in the control group had their heads held in a midline position, and no contact was made with the infant's face.

Tests of responsiveness to laterally applied somesthetic stimulation of the perioral region were begun within 2 minutes after the end of the 15-minute pretest period. Prior to testing, all wrappings were removed and sleeping infants were awakened, but once testing was begun no further attempt was made to alter the infant's state.

Because the infant's response may vary with his head position at the moment when a lateralized stimulus is applied, the heads of all *S*s were held in the midline prior to and during the application of the test stimuli. The infant's head was held at the temporal regions between the *E*'s thumb and finger. Stimuli were not presented until the infant's head was in the midline and the *E* could detect no lateralized pressure against the restraining fingers.

A test stimulus consisted of touching the angle of the infant's mouth for 1 second with a no. 12 round camel's hair brush. The infant's head was released simultaneously with the removal of the brush. The head was repositioned to the midline before each stimulus. Stimuli were applied whether the infant was crying or not since previous experience had shown that there was no difference in responsiveness to such stimuli when the baby was and was not crying (Turkewitz *et al.*, 1965).

There was a minimum interval of 15 seconds between test stimuli. However, because of differences in the time required to obtain the necessary midline balance of the head, the interval was often longer than this minimum. As a result, the total time required for testing an infant ranged from 12 to 27 minutes with a mean test duration of 17 minutes. There were no differences among the four groups with respect to the mean time required to complete the test series.

Each *S* was tested with the same predetermined random order of stimuli. This consisted of three replicated sets of 15 presentations with each set containing ten somesthetic stimuli (five to the right and five to the left perioral region) and five mock stimuli (three to the right and two to the left). The mock presentations were identical to the stimulus presentations, except that during the mock trials the brush was merely moved towards the infant's cheek but did not touch it. The mock stimuli

were included in order to replicate the test conditions and sequence of stimulation of our earlier studies.

The response measure was the direction of the first lateral head movement which occurred in the 1-second interval following stimulation.

To prevent experimenter bias, the person who applied the stimulus, released the infant's head from its midline position, and noted the nature of the infant's response had no knowledge of the pretest group to which the infant belonged.

RESULTS

To determine the effect of asymmetry of tonus and of somesthetic prestimulation on responsiveness to subsequently applied somesthetic stimuli, the mean percentage of infants in the control group making head-turning responses was compared, across the 30 stimulus trials, with the mean percentage of infants in each of the experimental groups making such responses. As may be seen in Table 1, both asymmetrical tonus and

TABLE 1
EFFECT OF ASYMMETRICAL TONUS AND SOMESTHETIC PRESTIMULATION ON THE PERCENTAGE OF HEAD-TURNING RESPONSES

Group	% of test stimulus trials resulting in head-turning responses	
	Mean	SD
Control	44.8	10.50
Tonus	63.8	11.82
Somesthetic right	64.7	10.06
Somesthetic left	62.5	12.39

differential prestimulation of either side resulted in more head turns than occurred in the control group. Each of these differences was significant at less than the .01 level (Dunnett test, Edwards, 1963).

All of the experimental treatments resulted in an increase in ipsilateral head-turning responses, but none of the treatments resulted in a change in contralateral responses (Table 2). The mean percentage of trials resulting in ipsilateral responses was significantly greater in each experimental group than in the control group (p's $< .01$, Dunnett test). In contrast, the mean percentage of trials resulting in contralateral responses in each of the experimental groups did not differ from that in the control group (Dunnett test). Since contralateral responses were infrequent (Table 2) and were unaffected by either asymmetrical tonus or differential somesthetic prestimulation, all subsequent analyses were restricted to a consideration of ipsilateral responses.

To determine whether asymmetrical tonus and differential prestimula-

TABLE 2

EFFECT OF ASYMMETRICAL TONUS AND SOMESTHETIC PRESTIMULATION ON THE PERCENTAGE OF IPSILATERAL AND CONTRALATERAL HEAD-TURNING RESPONSES

Group		% of test stimulus trials resulting in	
		Ipsilateral responses	Contralateral responses
Control	Mean	37.3	7.5
	SD	11.76	5.09
Tonus	Mean	53.5	10.4
	SD	12.63	9.64
Somesthetic right	Mean	57.1	7.6
	SD	10.88	8.10
Somesthetic left	Mean	53.8	9.0
	SD	13.34	9.22

tion had increased ipsilateral head turning by increasing responsiveness to stimulation of one or both sides, the mean percentage of ipsilateral head turns in each of the experimental groups was compared with that in the control group for each side. As may be seen in Table 3, the percentage of ipsilateral responses to stimulation of each side was greater in all of the experimental groups than in the control group. For both the somesthetic right and somesthetic left groups, these differences were highly significant (p's $< .01$, Dunnett test). For the tonus group, the increase in the mean percentage of responses to stimulation of the right from that in the control group was significant at less than the .01 level, and the increase in the mean percentage of responses to stimulation of the

TABLE 3

EFFECT OF ASYMMETRICAL TONUS AND SOMESTHETIC PRESTIMULATION ON LATERAL DIFFERENCES IN THE PERCENTAGE OF IPSILATERAL HEAD-TURNING RESPONSES

Group		% of trials resulting in ipsilateral responses to test stimulation	
		of the right	of the left
Control	Mean	37.7	37.0
	SD	9.74	13.79
Tonus	Mean	59.5	47.4
	SD	9.83	12.44
Somesthetic right	Mean	61.5	52.7
	SD	11.18	8.88
Somesthetic left	Mean	56.6	51.0
	SD	12.83	13.68

left from that in the control group approached the conventional level of statistical significance ($p < .10$).

Although asymmetrical tonus and differential prestimulation tended to increase ipsilateral responsiveness to both loci of stimulation, the tonus and somesthetic-right treatments had more of an effect on responsiveness to stimulation of the right than on responsiveness to stimulation of the left. This is evidenced by the finding of significantly more ipsilateral head turns to stimulation of the right than of the left in both the tonus and somesthetic right groups ($t = 3.12$, $p < .01$; and $t = 2.68$, $p < .02$ in the tonus and somesthetic right groups, respectively) and no such lateral difference in the control group ($t = .14$). The somesthetic-left treatment not only did not have more of an effect on responsiveness to stimulation of the left than on responsiveness to stimulation of the right, but resulted in slightly (not significantly) greater responsiveness to right than to left stimulation (Table 3).

The results thus far presented indicate that asymmetry of tonus and of somesthetic prestimulation increases responsiveness to both right and left stimulation and produces lateral differences in responding. For our sample of infants, these findings are highly reliable and can be generalized to other randomly selected samples of 30 stimulus trials. However, since the mode of analysis utilized does not take into account the variability of responsiveness among infants, it does not enable an assessment of the likelihood with which similar effects would be obtained in other randomly selected groups of infants comparably tested. We therefore evaluated the effect of the experimental treatments in relation to the interindividual variability of responsiveness.

As may be seen in Table 4, infants in each of the experimental groups made more head-turning responses (ipsilateral and contralateral combined) than did infants in the control group. However, because of the high variability among infants in all of the groups, none of the differences between the control and the experimental groups was statistically sig-

TABLE 4

EFFECT OF ASYMMETRICAL TONUS AND SOMESTHETIC PRESTIMULATION ON THE PERCENTAGE OF HEAD-TURNING RESPONSES

Group	N	% of head-turning responses	
		Mean	SD
Control	13	44.8	27.35
Tonus	14	63.8	25.08
Somesthetic right	14	64.7	26.90
Somesthetic left	15	62.5	24.16

nificant (Dunnett test). Similarly, although infants in each of the experimental groups made more ipsilateral responses to both right and left stimulation than did infants in the control group (Table 5), only the increases in response to stimulation of the right in the tonus and right somesthetic groups approached significance ($p < .10$, Dunnett tests). In both the tonus and right somesthetic groups, the increase in ipsilateral responsiveness to stimulation of the right was greater than that to stimulation of the left as evidenced by the finding of significantly more ipsilateral responses to right than to left stimulation in the tonus group ($t = 2.33$, $p < .05$) and a tendency in this direction in the somesthetic right group ($t = 1.77$, $p < .10$). In the somesthetic left and control groups, no lateral differences in ipsilateral responding were found. Thus, other randomly selected groups of infants treated as were the infants in the tonus and somesthetic right groups are likely to exhibit a lateral difference in ipsilateral head-turning responses whereas infants treated as were those in the somesthetic left group are not likely to exhibit a lateral difference.

Since it was possible that any effect of the experimental treatments on responsiveness to somesthetic stimulation could have been diminished or enhanced with time, the results from the first and the third blocks of test trials were compared with respect to lateral differences in responding. There was no indication of a time trend inasmuch as the data from the third block of trials were not found to differ from those of the first block of trials for any of the groups (t tests for dependent samples).

In view of the small number of male infants in each of the four groups, no meaningful analysis of sex differences in the effects of the various experimental conditions was possible. However, in none of the groups

TABLE 5

EFFECT OF ASYMMETRICAL TONUS AND SOMESTHETIC PRESTIMULATION ON THE PERCENTAGE OF IPSILATERAL HEAD-TURNING RESPONSES

Group		% of ipsilateral head-turning responses to test stimulation		
		of both sides	of the right	of the left
Control	Mean	37.3	37.7	37.0
	SD	25.69	27.6	26.7
Tonus	Mean	53.5	59.5	47.4
	SD	26.11	28.1	25.6
Somesthetic right	Mean	57.1	61.5	52.7
	SD	30.79	33.1	30.5
Somesthetic left	Mean	53.8	56.6	51.0
	SD	26.30	24.4	29.4

were the data from any of the male *Ss* markedly different from those of the female *Ss*.

DISCUSSION

The results of the present study indicate that asymmetry either of muscle tonus or of somesthetic prestimulation produces a general increase in head-turning responses to subsequent somesthetic stimulation which is a reflection of increases in ipsilateral responses to stimulation of both the right and the left cheek. Moreover, asymmetry of muscle tonus and prestimulation of the right cheek result in a lateral difference in ipsilateral responding. That this lateral difference is due to a greater effect of the tonus and somesthetic right treatments on the side primed than on the other side is evidenced by the finding that both asymmetrical tonus and prestimulation of the right cheek resulted in significantly more ipsilateral head turns to stimulation of the right than to stimulation of the left whereas when neither of these factors were present (in the control group), no systematic lateral difference in ipsilateral responding was found.

Although the results indicate that both the tonus and the somesthetic right treatments affected the subsequent lateral differentiation of the infants studied, the relatively high interindividual differences among babies raises the question of the likelihood with which the same lateral differences would be found in other randomly selected groups of infants comparably treated. The finding of a significant difference in the frequency of ipsilateral responses to stimulation of the right and of the left in the tonus group and of a tendency ($p < .10$) in this direction in the somesthetic right group even when interindividual differences were taken into account, suggests that similar differences are likely to be obtained in other randomly selected samples of infants comparably treated. However, since in neither group were all of the infants more responsive to stimulation of the right than of the left, the existence of two types of infants is suggested, those whose lateral differentiation is, and those whose lateral differentiation is not affected by treatment.

In view of these findings, it seems likely that both asymmetrical muscle tonus and prestimulation of the right cheek contribute to the normally occurring lateral differences in ipsilateral responsiveness although both need not operate in all infants and indeed there might be some infants (or circumstances) in which neither is operative.

To determine whether the effects of asymmetry of muscle tonus and of prestimulation of the right cheek must summate or interact with each other to produce the degree of lateral differentiation which occurs when both factors are simultaneously present, the results of the current investigation were compared with those of an earlier investigation (Turke-

witz *et al.*, 1967) in which the infant's spontaneous head-right posture was not interfered with. Since the mean percentage of lateral differences obtained following the latter condition in which both factors were present was not significantly different (t tests) than the mean percentages of lateral differences when only one or the other was present, there is no indication that the effects of asymmetry of muscle tonus and of somesthetic prestimulation of the right cheek either summate or interact with each other.

The effects of prior somesthetic stimulation on lateral differentiation cannot be accounted for in terms of a simple, direct effect of immediately preceding input but must be viewed as a consequence of the interaction of the effects of lateral differences in immediately preceding input and the effects of another lateral difference in organization. If prior stimulation of one side were sufficient to produce greater responsiveness to subsequent stimulation of that side than of the other side, infants prestimulated at the right side would have exhibited more ipsilateral responses to stimulation of the right than of the left, and infants prestimulated at the left would have exhibited greater responsiveness to stimulation of the left than of the right. Since prestimulation of the right had this effect but prestimulation of the left did not, and in fact resulted in slightly more ipsilateral responses to stimulation of the right than of the left, the hypothesis of a single determinative factor must be rejected. It is likely that the effects of prior stimulation of the right side summate with the effects of another factor which increases responsiveness on the right to produce greater responsiveness to stimulation of the right than of the left side. The effects of prior stimulation of the left, on the other hand, are insufficient to counteract the effects of this factor which increases responsiveness at the right, and so a tendency for greater responsiveness to stimulation of the right than of the left is still manifested. It remains for future investigations to identify and determine the basis of this factor which affects lateral differences in responsiveness.

Any such future investigation could focus on the identification of structural asymmetries which may be directly determined by genetic factors, on intrauterine conditions, or on the continuing effects of the prior operation of precisely those factors of prestimulation with which we have been concerned.

REFERENCES

APGAR, V., & JAMES, L. S. Further observations on the newborn scoring system. *American Journal of Diseases of Children*, 1962, **104**, 419.

EDWARDS, A. L. *Experimental Design in Psychological Research*. New York: Holt, 1963.

GESELL, A., & AMES, L. B. The development of handedness. *Journal of Genetic Psychology,* 1947, **70,** 155–175.

SIQUELAND, E. Operant conditioning of head turning in four-month infants. *Psychonomic Science,* 1964, **1,** 223–224.

TURKEWITZ, G., BIRCH, H. G., MOREAU, T., LEVY, L., & CORNWELL, A. C. Effect of intensity of auditory stimulation on directional eye movements in the human neonate. *Animal Behaviour,* 1966, **14,** 93–101.

TURKEWITZ, G., GORDON, E. W., & BIRCH, H. G. Head turning in the human neonate: Effect of prandial condition and lateral preference. *Journal of Comparative and Physiological Psychology,* 1965, **59,** 189–192.

TURKEWITZ, G., MOREAU, T., & BIRCH, H. G. Head position and receptor organization in the human neonate. *Journal of Experimental Child Psychology,* 1966, **4,** 169–177.

TURKEWITZ, G., MOREAU, T., & BIRCH, H. G. Relation between birth condition and neuro-behavioral organization in the neonate. *Pediatric Research,* 1968, **2,** 243–249.

TURKEWITZ, G., MOREAU, T., BIRCH, H. G., & CRYSTAL, D. Relationship between prior head position and lateral differences in responsiveness in the human neonate. *Journal of experimental child Psychology,* 1967, **5,** 548–561.

WICKELGREN, L. W. Convergence in the human newborn. *Journal of experimental child Psychology,* 1967, **5,** 74–85.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 494–502 (1969)

Transfer from Perceptual Pretraining as a Function of Number of Task Dimensions[1]

LOUISE S. TIGHE AND THOMAS J. TIGHE
Dartmouth College

Children (average age 6 years 7 months) learned an initial discrimination (OL) and a reversal shift (RS) with 1 dimension relevant and either 0, 1, or 2 dimensions irrelevant following pretraining which required *S*s to make same-different judgments to stimulus objects varying along the dimensions appearing in the transfer tasks. Pretraining did not affect performance in OL relative to a control condition, and performance degraded as a linear function of number of irrelevant dimensions. In RS, perceptually pretrained and control *S*s did not differ under the condition of 0 dimensions irrelevant, but with increasing amounts of irrelevant stimulation the performance of the control *S*s deteriorated in the same manner as in OL while the performance of perceptually pretrained *S*s was unaltered.

Previous experiments have defined a set of pretraining conditions of a perceptual nature which produce facilitation of discrimination reversal learning in 5–7-year-old children (Tighe, 1965; Tighe & Tighe, 1968a). The major feature of the pretraining is that *S*s are required to make nonreinforced same-different judgments to successively presented stimulus objects which vary along the dimensions appearing in the subsequent discrimination tasks. Following the conceptions of differentiation theory (Gibson & Gibson, 1955; Tighe & Tighe, 1966), which provided the impetus for these experiments, the observed facilitation is deduced from the assumption that perceptual pretraining increases *S*'s ability to isolate and independently utilize the task dimensions. In contrast, nonpretrained *S*s of this age are assumed to be more likely to accomplish the initial discrimination and reversal on the basis of the specific stimulus object-reward relations in each problem. Only when *S* has related the relevant dimension, rather than the specific objects, to reward is there a basis for

[1] This research was supported by Public Health Service Research Grant MH 11088 from the National Institute of Mental Health. The authors are grateful for the generous cooperation of Gordon Tate, Superintendent of the Lebanon, New Hampshire, school system, and John A. Freitas, Superintendent of the Hartford, Vermont, school system, and for the assistance of the principals and teachers of these school systems. We also wish to acknowledge the contribution of Mrs. Jane Higgens, who assisted in collecting the data.

positive transfer to a subsequent reversal shift in the form of attention to the relevant feature-reward relation. This analysis implies that the beneficial effects of perceptual pretraining should become increasingly evident with increases in the number of nonredundant task dimensions since arithmetic increase in number of dimensions requires geometric increase in the number of stimulus objects necessary to represent the possible variations. The present experiment studies this issue by comparing the performance of perceptually pretrained and nonpretrained children on transfer tasks with 1 relevant and 0, 1, or 2 irrelevant dimensions.

METHOD

Subjects. The *S*s were 252 first-grade children average age 6 years and 7 months from the Hartford, Vermont, and the Lebanon, New Hampshire, school systems. Each *S* was run individually in a small room adjacent to his classroom.

Pretraining. The stimuli consisted of wood blocks which varied in three values of height (Ht), brightness (Br), and shape (Sh). Brightness was varied by covering the blocks with black (B) or white (W) enamel paint or with a medium gray (1 part B to 3 parts W). The Ht values were 4¾ inches, 5½ inches, and 6¼ inches, and the shapes were cylindrical (1½-inch diameter), square (1⅝-inch sides), and hexagonal (1¾ inches at maximum width). The stimuli were placed one at a time about 15 inches from *S* who was seated opposite *E* at a small table. Between presentations all stimuli were hidden from *S*'s view behind a nearby masonite screen.

Half of the *S*s were perceptually pretrained, and these *S*s were further subdivided into three equal groups for training with either 1, 2, or 3 variable dimensions. For *S*s pretrained with variation in 3 dimensions, the pretraining stimuli consisted of the 27 possible combinations of the Ht, Br, and Sh values described above. The total presentation was divided into 8 subsets or series as described below. The first stimulus presented within a series constituted a standard (St) stimulus. The *S* was told, "Now let's pretend that this object is yours. Look at it carefully. I am going to take it away and show you some other objects one at a time. You tell me which of the other objects that you will see are exactly like yours." As each succeeding comparison (Co) stimulus was placed in the stand, *E* said, "Is this object exactly like yours?", and the stimulus remained until *S* had rendered a judgment of "same" or "different." When an St was presented again to begin a new series, the first sentence of the instructions was changed to, "Here is your object again." When a new St was introduced, the first sentence of the instructions was, "Now let's pretend that this object and only this one is yours." Eight St stimuli

were used, consisting of the eight possible combinations of the 4¾-inch and 6¼-inch heights, the square and cylindrical shapes, and the B and W brightness values. Subjects underwent one series of Co stimulus presentations with each of the eight Sts. A series included the 27 possible combinations of the three Ht, Br, and Sh values, plus from two to three interspersed reappearances of the St stimulus.

Subjects perceptually pretrained with either 2-dimensional or 1-dimensional variation in the stimulus objects received training identical to that given *Ss* trained with variation in three dimensions with the exception of the stimulus objects constituting the judgment series. For *Ss* trained with 2-dimensional variation, four different Sts were used. The stimulus values specifying these Sts were the same as those comprising the Sts used in pretraining with 3-dimensional variation, but were selected so as to be constant in one dimension. For example, for a given *S* the four Sts might consist of the possible combinations of the 4¾-inch and 6¼-inch heights and the B and W brightness values but with all stimuli hexagonal in shape. Across *Ss*, each dimension was variable and held constant equally often. An *S* underwent two series of Co stimulus presentations with each of the four Sts designated for him. A series comprised the nine possible combinations of the three values of the dimensions allowed to vary in the representation of the St stimuli plus from one to two interspersed reappearances of the St. The dimension held constant in St stimulus representation was also held constant during the Co series and in the same manner, i.e., the same value of that dimension appeared during the St and Co presentation series. For the Ht and Br dimensions, this was the intermediate value, while for the Sh feature it was the hexagonal shape which appeared.

For *Ss* trained with 1-dimensional variation, two Sts were used. Again, the stimulus values specifying these Sts were the same as those comprising the Sts used in pretraining with 3-dimensional variation, but were selected so as to be constant in two of the dimensions. For example, for a given *S* the two Sts might consist of the 4¾-inch and 6¼-inch stimuli identical in Br and Sh. Across *Ss*, each dimension was variable and held constant equally often. An *S* underwent four series of Co stimulus presentations with each of the two Sts designated for him. A series included the three values of the varying dimension plus one to two interspersed reappearances of the St. The dimensions held constant during the presentation of the St stimuli were also held constant during the Co series, as described above in connection with the 2-dimensional variation condition.

The possible orders of the Sts appearing in each pretraining condition were randomly distributed among *Ss* with the restriction that for half the *Ss* the last St presented would be identical to one of the positive

stimuli in the later discrimination learning and for the other half the last St would be identical to one of the negative stimuli. None of the judgments were reinforced or corrected and all were based solely upon vision. The number of experiences with each Co stimulus was the same for all conditions of the experiment—eight experiences per stimulus. The time required to complete pretraining varied from about five minutes in the 1-dimensional variation condition to approximately 20 minutes in the 3-dimensional variation condition.

The Ss who did not receive perceptual pretraining ($N = 126$) worked on tasks unrelated to the subsequent discrimination problems and designed to control for nonspecific transfer effects. The tasks were modifications of the Picture Completion and Picture Arrangement subtests of the Wechsler Intelligence Scale for Children. The modifications consisted of (*a*) presenting all 20 pictures in the completion test to each *S* and (*b*) requiring each *S* to tell a story about each picture in the arrangement test. None of *S*'s responses in either task were corrected or reinforced. Training on these tasks was carried out for three time periods corresponding to the mean times required to perceptually pretrain *S*s with 1, 2, or 3 variable dimensions. An equal number of control *S*s were trained for each time period.

Discrimination tasks. Immediately following pretraining all *S*s learned a two-choice discrimination and a discrimination reversal involving the dimensions appearing in pretraining. For perceptually pretrained *S*s, these transfer tasks contained the same variable and constant dimensions as those appearing during *S*'s pretraining. Thus, *S*s having only one dimension variable during pretraining had only that same dimension variable and relevant during discrimination learning. Subjects who had two dimensions varying during pretraining had these dimensions varying during subsequent learning, with one relevant and one irrelevant to reward; and *S*s pretrained with three variable dimensions had all features varying during learning with one relevant and two irrelevant. Dimensions held constant during learning were represented by the same stimulus value used to represent the constant dimensions during pretraining. The stimuli were presented on a simple turntable device of the type described by Kendler and Kendler (1959). For *S*s trained with one, two, or three dimensions variable, one, two, or four pairs of choice objects, respectively, were required to represent the possible variations; and these choice objects were the St stimuli appearing in *S*'s pretraining. An equal number of control *S*s, matched to the experimental *S*s in pretraining time, learned the discrimination and reversal tasks under the three conditions of dimensional representation.

In each learning condition all dimensions were relevant equally often,

and both values of each dimension were used equally often as the S+ and S—. The *S* was instructed to pick up one of the objects on each trial and was told that if his choice were correct he would find a marble under it. The *S* was told that the "game" was to see how soon he could find a marble every time he chose. Training was continued in this manner until *S* made nine correct out of ten successive responses, at which point without change of instruction the discrimination reversal was initiated and carried to the same criterion. Manner of dimensional representation was not altered during reversal, i.e., the irrelevant dimension(s) continued to vary as in the initial discrimination. Throughout training the stimuli were presented according to a prearranged sequence designed to control for order and position effects. Each *S* was assigned to groups and to positive stimuli on a predetermined random basis. All *S*s trained attained criterion in the initial discrimination and reversal. At the end of the experiment, *S* was allowed to choose a prize from an assortment which included charms, bubble gum, whistles, raisins, jack sets, Tootsie rolls, pencils, pencil sharpeners, M & Ms, five pennies, flutes, and crayons.

RESULTS

The mean number of errors during pretraining on the first and last series of judgments were .48 and 0, 1.29 and .26, and 3.10 and .83 for *S*s pretrained with 1, 2, and 3 variable dimensions, respectively. By dependent t test, fewer errors were made on the final judgment series as compared to the first series in each pretraining condition. For *S*s trained with 1, 2, and 3 dimensions variable, $t = 5.58$, 7.40, and 4.13, respectively; $df = 41$ in each case, and all p values $< .01$.

Table 1 presents the mean number of trials and errors to criterion in the original learning (OL) and in reversal shift (RS) for the major treatment groups. Factorial analyses of variance ($2 \times 3 \times 3$) were carried out on $\sqrt{X + .5}$ transformed trials to criterion in both tasks. For the

TABLE 1

MEAN NUMBER OF TRIALS AND ERRORS TO CRITERION IN THE ORIGINAL LEARNING (OL) AND REVERSAL (R) FOR THE MAJOR TREATMENT GROUPS

Task	Type of pretraining	Number of irrelevant dimensions: 0 Trials	0 Errors	1 Trials	1 Errors	2 Trials	2 Errors
OL	Perceptual	12.0	6.2	21.9	10.5	31.6	15.9
	Control	9.2	4.8	26.4	13.0	29.4	14.9
R	Perceptual	4.7	2.7	5.3	3.3	4.7	3.2
	Control	6.8	3.4	26.3	13.9	42.9	21.6

OL, the analysis revealed significant main effects for both number of irrelevant dimensions ($F = 29.8$, $df = 2,234$, $p < .001$) and for type of dimension relevant ($F = 3.89$, $df = 2,234$, $p < .025$). Type of pretraining (i.e., perceptual vs. control pretraining) did not approach significance ($F < 1$), nor were any of the interactions a significant source of variance. Trend analysis of the decline in learning speed with increase in the number of irrelevant dimensions was carried out on the nontransformed scores. This was deemed appropriate since factorial analyses of variance of transformed and nontransformed data yielded identical conclusions. The trend analysis of the performance-irrelevant dimension relation indicated a significant linear component only ($F = 65.5$, $df = 1,249$, $p < .001$).

In further analysis of the type of dimension variable, the data were combined over the remaining treatments and t tests were employed to compare speed of learning with the different relevant dimensions. Learning with Ht relevant was more rapid than learning with Br relevant ($t = 2.10$, $df = 166$, $p < .05$) or than learning with Sh relevant ($t = 2.72$, $df = 166$, $p < .01$). Learning speed did not differ under Br relevant vs. Sh relevant ($t = .50$, $df = 166$, $p > .60$). (Probabilities are 2-tailed.)

Factorial analysis for RS learning revealed significant main effects for both type of pretraining ($F = 157.4$, $df = 1,234$, $p < .001$) and number of irrelevant dimensions ($F = 31.7$, $df = 2,234$, $p < .001$), and a significant interaction between these factors ($F = 62.0$, $df = 1,234$, $p < .001$). Type of dimension relevant was not a significant condition of RS learning ($F = 1.72$, $df = 2,234$, $p > .10$), and no other interaction effects were significant. Trend analysis of the performance-number of irrelevant dimension function for the control *Ss* indicated a significant linear component only ($F = 75.6$, $df = 1,123$, $p < .001$). A significant trend did not appear in this relation for the perceptually pretrained *Ss* ($F < 1$). Duncan's multiple comparison technique was employed to compare learning speed among the perceptually pretrained and control groups learning under the conditions of 0, 1, or 2 irrelevant dimensions (designated as Groups P0, P1, P2, and C0, C1, and C2, respectively). Groups C0, P0, P1, and P2 did not differ from one another ($p > .10$), but all differed from Group C1 ($p < .001$) which, in turn, differed from Group C2 ($p < .001$).

Analysis of errors to criterion in OL and RS produced outcomes identical to those described for the trials to criterion measure.

To summarize, then, perceptual pretraining did not affect performance in OL relative to the control conditions and performance degraded as a linear function of number of irrelevant dimensions. In RS learning, perceptually pretrained and control *Ss* did not differ under the condition

of 0 dimensions irrelevant, but with increasing amounts of irrelevancy the performance of control *Ss* deteriorated in the same manner as in OL while the performance of perceptually pretrained *Ss* was unaltered.

DISCUSSION

The fact that the RS performance of the pretrained *Ss* was not disturbed by increase in the number of irrelevant dimensions indicates that discriminative response in these *Ss* was firmly under the control of the relevant dimension at the start of RS. From the differentiation viewpoint (Tighe & Tighe, 1966; 1968b), this control resulted primarily from an increase, through perceptual pretraining, in *S*'s ability to detect the separate dimension-reward relations of the discrimination tasks, and secondarily from the development of attention to the relevant dimension during OL. When *S* reacts selectively to the relevant dimension in RS, the reversed relation between the values of this feature and reward is likely to be readily detected within an otherwise constant stimulus milieu, regardless of the number of irrelevant features. In contrast, the progressive deterioration in the RS learning of nonpretrained *Ss* with increase in irrelevant dimensions is seen as reflecting the tendency of these *Ss* to detect only object-reward relations in the tasks. Such an inferior mode of discrimination might take several forms. One possibility which we have suggested elsewhere (Tighe & Tighe, 1966) is that *Ss* may relate reward or nonreward to each object perceived as a whole. However, it seems unlikely that the nonpretrained *Ss* of the present experiment discriminated in this fashion since the relation between learning speed and number of dimensions does not approach the positively accelerated form which would be expected under such learning. Another possibility is that the nonpretrained *Ss* respond in OL and RS to the combined cue properties of the objects with response to the S+ and S− cues becoming predominant as training progresses. Under this view the decline in learning speed with increase in dimensions follows directly from the decline in the proportion of relevant to irrelevant cues with increase in dimensions.

The fact that there was no difference in the RS learning of pretrained and nonpretrained *Ss* under the condition of 0 dimensions irrelevant is not surprising from the differentiation viewpoint. As detailed in Tighe and Tighe (1966), the procedures of perceptual pretraining were designed to increase *S*'s ability to utilize the distinguishing features of the task stimuli *independently* of one another, and such perceptual learning is assumed to be the major benefit of pretraining. Consequently, there is little reason to expect advantage from pretraining when only one dimension appears in the transfer tasks. It is true, of course, that spatial position of S+ and S− is present as an irrelevant feature in the 0

dimensions irrelevant condition. But the relation of this feature to the relevant dimension does not have the "embedded" quality which obtains in the relations between the various visual dimensions.

Attentional theories of discrimination learning (e.g., Sutherland, 1959; Zeaman & House, 1963), which relate speed of reversal to the strength of *S*'s attention to the relevant dimension, might deduce the observed facilitation of RS from the assumption that perceptual pretraining serves to focus *S*'s attention on the task dimensions vis-à-vis other features in the general experimental situation. As a consequence of this altered attention, pretained *S*s would be likely to have a higher probability of attending to the relevant dimension at the start of RS than *S*s who undergo only OL. Under this view, the failure to find a difference between the RS learning of pretrained and nonpretrained *S*s under the condition of 0 dimensions irrelevant indicates that this condition was especially favorable to the development of attention to the relevant dimension during OL itself.

In regard to the OL, the observed linear relation between number of irrelevant dimensions and rate of learning is consistent with the outcomes of a number of studies involving similar discrimination tasks (Bourne & Restle, 1959). But the fact that pretraining did not facilitate performance under the higher degrees of irrelevant stimulation is surprising from the viewpoint of both differentiation and attention theory, since the altered sensitivity to the relevant dimension which is assumed to underlie facilitation of RS should also obtain during the initial task. In discussing previous failures to observe an effect of perceptual pretraining in 2-dimensional discriminations (Tighe & Tighe, 1968a), it was noted that the OL, unlike reversal, requires considerable nonspecific learning (e.g., the elimination of position or sequence hypotheses) as well as the detection of the relevant dimension, and it was suggested that pretraining probably did not constitute an appreciable advantage in the face of such nonspecific learning. This argument is weakened somewhat by the present observation that pretraining also does not affect performance in a 3-dimensional OL problem, since this task presumably represents a significant increase over a 2-dimensional task in specific stimulus-reward learning relative to nonspecific learning requirements. However, a more rigorous test of the hypothesis should be made. One possibility would be to give *S*s, prior to perceptual pretraining, a series of discriminations involving dimensions other than those appearing in pretraining. In this manner, the *S*s should begin the discrimination following pretraining having mastered the "rules of the game," and their learning should be primarily concerned with detection of the feature-reward relations specific to the OL. Such *S*s should learn the discrimination following

perceptual pretraining more rapidly than *Ss* who receive the prior discrimination tasks but who are not perceptually pretrained.

REFERENCES

BOURNE, L. E., JR., & RESTLE, F. A mathematical theory of concept identification. *Psychological Review,* 1959, **66,** 278–296.

GIBSON, J. J., & GIBSON, E. J. Perceptual learning: Differentiation or enrichment? *Psychological Review,* 1955, **62,** 32–41.

KENDLER, T. S., & KENDLER, H. H. Reversal and nonreversal shifts in kindergarten children. *Journal of Experimental Psychology,* 1959, **58,** 56–60.

SUTHERLAND, N. S. Stimulus analysing mechanisms. *In* "Proceedings of a Symposium on the Mechanisation of Thought Processes," Vol. 2. London: Her Majesty's Stationary Office, 1959. Pp. 575–609.

TIGHE, L. S. Effect of perceptual pretraining on reversal and nonreversal shifts. *Journal of Experimental Psychology,* 1965, **70,** 379–385.

TIGHE, L. S., & TIGHE, T. J. Discrimination learning: Two views in historical perspective. *Psychological Bulletin,* 1966, **66,** 353–370.

TIGHE, T. J., & TIGHE, L. S. Perceptual learning in the discrimination processes of children: An analysis of five variables in perceptual pretraining. *Journal of Experimental Psychology,* 1968, **77,** 125–134. (a)

TIGHE, T. J., & TIGHE, L. S. Differentiation theory and concept-shift behavior. *Psychological Bulletin,* 1968, **70,** 756–761. (b)

ZEAMAN, D., & HOUSE, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of Mental Deficiency.* New York: McGraw-Hill, 1963. Pp. 159–223.

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 8, 503–511 (1969)

Cue Novelty and Training Level in the Discrimination Shift Performance of Retardates[1]

LINDA BILSKY[2] AND LAIRD W. HEAL[3]

George Peabody College for Teachers

This study assessed the effects of cue novelty and training level on the attention of moderately retarded adults. Each *S* learned a problem, to one of three training levels, with his nonpreferred dimension relevant and his preferred dimension irrelevant. A subsequent shift problem, employing either familiar or novel cues on *S*'s preferred dimension, allowed two alternative solutions, (a) a reversal based on *S*'s nonpreferred dimension or, (b) an extradimensional shift based on *S*'s preferred dimension. *S*s used more extradimensional shift solutions in the presence of novel cues than in the presence of familiar cues. This effect was absent at the highest training level.

Recently, the role of attention in the discrimination learning of mentally retarded individuals has become an issue of considerable importance. Zeaman & House (1963, p. 188) have proposed that "retardates suffer from low initial probability of observing certain relevant dimensions rather than from poor ability to learn which of two observed cues is correct." Consequently, it may be possible to improve the discrimination performance of mentally retarded individuals through the manipulation of factors influencing attention. The present study investigated two of these factors, (a) cue novelty and (b) training level.

The introduction of novel stimulus materials has facilitated performance in a variety of situations. Cantor and Cantor have repeatedly observed faster response speeds and longer fixation times for novel, as opposed to familiar, stimulus lights and pictures (G. Cantor & J. Cantor,

[1] This paper is based on a Master's thesis submitted to the Graduate School, George Peabody College for Teachers, June 1967. An earlier version of this manuscript was read at the annual convention of the American Association on Mental Deficiency, Denver, May 1967. The study was supported by NICHD Grants HD-43 and HD-973.

[2] Currently at Teachers College, Columbia University.

[3] The authors wish to thank the staff at Clover Bottom Hospital and School, Donelson, Tennessee, for their cooperation in furnishing subjects for this study. The authors are also grateful to Dr. Leonard Blackman and Dr. Ross Evans for their helpful criticism of the manuscript.

1964, 1965, 1966; J. Cantor & G. Cantor, 1964a, 1964b). Other investigators (Zeaman, House, & Orlando, 1958) have reported that the introduction of either positive or negative novel cues significantly improved the performance of retarded *Ss* who had failed to learn a discrimination problem within 250 trials. Guy, Van Fleet, & Bourne (1966) found better extradimensional shift performance when a novel relevant dimension was introduced at the time of shift than when this dimension was present, but irrelevant, during the preshift problem. These and similar findings suggest that novel stimulus materials tend to elicit *S*'s attention more readily and to sustain his attention for longer periods of time than familiar stimulus materials.

Another factor of major interest in the present study was training level. It has been observed repeatedly that large amounts of training on a discrimination problem tend to facilitate performance on subsequent reversals of that problem. This phenomenon, known as the overtraining reversal effect (ORE), has been adequately reviewed elsewhere (Wolff, 1967) and is probably restricted to discrimination transfer situations which require *S* to attend to the same dimension in both problems. Large amounts of training generally have not facilitated extradimensional (ED) shifts which require *S* to direct his attention to a new relevant dimension.

The present study was an attempt to explore the effects of cue novelty and training level on *Ss*' tendencies to shift attention from one dimension to another in the solution of discrimination problems. Two hypotheses stated that (a) tendencies to shift attention would be greater with novel cues than with familiar cues and (b) the effects of novelty would be observed at low levels of training but not at high levels of training.

The effects of cue novelty and training level were assessed in a $2 \times 2 \times 3$ factorial design. The first factor, cue novelty, specified that on a shift problem (Stage 3) half the subjects would receive novel cues and half would receive familiar cues from a dimension which had been irrelevant during a preshift problem (Stage 2). The second factor, dimension, specified that for half the *Ss* in each cue novelty group color would be relevant during Stage 2 and for the rest of the *Ss* form would be relevant during Stage 2. The third factor, training level, specified that equal numbers of *Ss* would receive each of three levels of training on Stage 2.

Heal, George, & Bransky (1965) noted that dimension preference was an important factor in shifts of attention from one dimension to another. They observed that *Ss* tended to shift attention to novel cues on their preferred dimensions, but not to novel cues on their nonpreferred dimensions. Consequently, each *S* in the current study was assigned to a preshift problem (Stage 2) in which his nonpreferred dimension was relevant and his preferred dimension was irrelevant.

METHOD

Subjects

The *S*s were 72 institutionalized moderately retarded young adults with a mean CA of 24–10 and a mean IQ of 47.9. Equal numbers of color-preferring and form-preferring *S*s were selected on the basis of dimension preference information from previous discrimination studies in which they had participated. Twenty-four *S*s were eliminated because of illness or failure to reach the Stage 2 criterion. Each of the remaining 48 *S*s was assigned to the dimension condition in which his nonpreferred dimension was relevant during Stage 2. Equal numbers of *S*s from each of these groups were randomly assigned to cue novelty and training level conditions.

Materials

A two-choice discrimination apparatus was fitted with a pair of Industrial Electronics readouts (80-014R-P). These readouts projected 2″ × 2″ forms onto 2″ × 4″ colored backgrounds with inner edges 6″ apart. Both projectors were turned off when either panel was pressed. On feedback trials, the sound of a door chime indicated a correct response and the sound of a buzzer indicated an error.

The stimuli, which consisted of several different color-form combinations, are described in Table 1. Each of two cues on the irrelevant dimension was paired with the correct relevant cue on 50% of the trials and with the incorrect relevant cue on the other 50%. Position was counterbalanced and cues were presented in the same unsystematic cyclic 24-trial sequence to all subjects.

Procedure

At the beginning of his experimental session each *S* was instructed always to press the panel that rang the bell and never the buzzer. He was also told that although he would hear nothing on some trials, he should keep responding in the same way. Feedback was withheld on 33% of training trials and 100% of Stage 3 test trials.

Each session consisted of three stages. Essentially, Stage 1 was a pretesting phase which was included to allow the reassessment of dimension preference on the first trial of Stage 2. During Stage 1, a specific color-form compound was rewarded. As soon as *S* made four consecutive correct responses on feedback trials, he was presented with Stage 2.

Trial 1 of Stage 2 provided a brief reassessment of *S*'s dimension preference. Stage 1 color and form cues were re-paired on the first trial of Stage 2 forcing *S* to choose between the previously correct color and the

TABLE 1
EXPERIMENTAL CONDITIONS AND STIMULI[a]

	Form preferring *Ss*				Color preferring *Ss*			
	Novel forms		Familiar forms		Novel colors		Familiar colors	
	+	−	+	−	+	−	+	−
Stage 1	C_1F_1	C_2F_2	C_1F_4	C_2F_3	F_5C_3	F_6C_4	F_5C_6	F_6C_5
	+	−	+	−	+	−	+	−
Stage 2	C_1F_2	C_2F_1	C_1F_3	C_2F_4	F_5C_4	F_6C_3	F_5C_5	F_6C_6
	+	−	+	−	+	−	+	−
	C_1F_1	C_2F_2	C_1F_4	C_2F_3	F_5C_3	F_6C_4	F_5C_6	F_6C_5
	+	−	+	−	+	−	+	−
(Special training)	C_1	C_2	C_1	C_2	F_5	F_6	F_5	F_6
	+	−	+	−	+	−	+	−
Stage 3	C_2F_3	C_1F_4	C_2F_3	C_1F_4	F_6C_5	F_5C_6	F_6C_5	F_5C_6
	0	0	0	0	0	0	0	0
Choice trials	C_2F_4	C_1F_3	C_2F_4	C_1F_3	F_6C_6	F_5C_5	F_6C_6	F_5C_5
	C_1 white				F_1 square			
	C_2 yellow-blue				F_2 square with superimposed cross			
	C_3 red-amber				F_3 square with superimposed circle			
	C_4 green				F_4 star			
	C_5 yellow				F_5 "X"			
	C_6 red				F_6 circle with superimposed "X"[b]			

[a] In order to aid in the interpretation of Table 1, in which numerically subscripted letters represent cues and dimensions, the following example is provided: For form preferring *Ss* in the novel form condition, a white square was the positive cue (C_1F_1) and a yellow-blue square with a superimposed cross was the negative cue (C_2F_2) during Stage 1.

[b] Due to an apparatus failure, 3 form preferring *Ss* in the Novelty group and 8 form preferring *Ss* in the Familiarity group were presented with a circle with a superimposed star rather than a circle with a superimposed "X". These *Ss* were retained in their assigned conditions.

previously correct form. His choice on this trial was considered an indication of his dimension preference. During Stage 2 only a response based on *S*'s nonpreferred dimension was correct. Consequently, most *Ss* were expected to make errors on trial 1.

The remaining trials of Stage 2 constituted the preshift training phase of the present study. During Stage 2, each *S* was trained to respond on the basis of his nonpreferred dimension and to ignore his preferred dimension. His nonpreferred dimension was relevant and his preferred dimension was irrelevant. Thus, if a green square had been rewarded in

Stage 1, a color-dominant *S* would have had square rewarded in Stage 2 regardless of the color with which it was paired. The effect of dimension preference was supplemented by employing fairly obvious cues on *S*'s preferred dimension and less obvious cues on his nonpreferred dimension.

The Stage 2 learning criterion depended upon the training level to which *S* had been assigned and consisted of either 4, 8, or 16 consecutively correct responses on feedback trials with no errors on intervening non-feedback trials. If *S* failed to reach his learning criterion within 24 trials, special training was immediately introduced. Special training consisted of training with the relevant cues only (irrelevant cues were eliminated) to a criterion of 8 consecutive correct responses. If *S* attained this criterion, Stage 2 was repeated. If he again failed to reach the Stage 2 criterion, special training was given once more followed by a third presentation of Stage 2. Training was terminated at the first error following the 24th trial if *S* failed to reach criterion either during special training or his third attempt at Stage 2.

Stage 3 was a shift problem consisting of training trials on which a particular color-form compound was always rewarded and test trials on which neither response was rewarded. On their nonpreferred dimension, all *S*s received cues which had been present during Stages 1 and 2. However, on their preferred dimension, half the *S*s received new cues and the other half received cues which had been present during Stages 1 and 2. On training trials, Stage 2 reward contingencies were reversed with one cue from the Stage 2 irrelevant (preferred) dimension consistently paired with the new negative cue and the other consistently paired with the new positive cue. Thus if square had been rewarded in Stage 2, the red circle would be rewarded in Stage 3. Seven non-feedback test trials were interspersed among Stage 3 training trials. One test trial occurred after each series of three consecutive rewarded responses. On test trials, cues were repaired so that *S* had to choose between the components of the color-form compound which had been rewarded on training trials. The purpose of the test trial was to reveal whether *S* continued to attend to the dimension which had been rewarded during Stage 2, or whether he shifted his attention to his preferred dimension. If *S* made reversal responses on test trials, it was surmised that he was continuing to attend to his nonpreferred dimension which had been relevant during preshift training (Stage 2). On the other hand, if *S* made ED shift responses on test trials, it was surmised that his attention had shifted to his preferred dimension which had been irrelevant during Stage 2. Because choice trials provided no feedback, reward itself could not be the basis of *S*'s choice of a reversal or an ED shift. After his session each *S* was given a nickel as a reward for participating in the experiment.

RESULTS

Stage 1

Experimental groups were compared in terms of number of errors on Stage 1 feedback trials. A $2 \times 2 \times 3$ factorial analysis of variance with these data failed to reveal significant differences among groups ($p > .05$ for all comparisons).

Stage 2

Stage 2 data were analyzed by $2 \times 2 \times 3$ analyses of variance on Trial 1 errors as well as on overall errors on feedback trials. Twenty-one of the 24 *Ss* assigned to the color problem made errors on Trial 1, the first trial on which color and form cues were re-paired. However, of the 24 *Ss* assigned to the form problem, only 14 made errors on Trial 1, which was significantly fewer than the number making errors on the color problem, $F(1,36) = 5.88$, $p < .05$. The color-relevant group also made more errors than the form-relevant group during the rest of Stage 2, $F(1,36) = 4.68$, $p < .05$. These findings supported the assumption that those *Ss* assigned to the color-relevant problem had been appropriately assigned to their nonpreferred dimension. However, the appropriateness of this assignment appeared less certain for those *Ss* assigned to the form-relevant problem.

The analysis of overall errors on Stage 2 feedback trials revealed a significant difference among training level groups, $F(2,36) = 3.71$, $p < .05$. Mean errors on feedback trials were 5.44 for the training-level-4 group, 5.62 for the training-level-8 group, and 2.25 for the training-level-16 group. Nine *Ss* each at training levels 8 and 16, but only 3 *Ss* at training level 4, were eliminated for failure to reach criterion. Both of these findings suggested that only the best learners were able to attain the more difficult criterion.

Stage 3

A $2 \times 2 \times 3$ factorial analysis of variance of errors on feedback trials during Stage 3 training failed to reveal any group differences. Test trial data were analyzed by a $2 \times 2 \times 3$ factorial analysis of variance on number of ED shift responses. Number of ED shifts was greater in the presence of novel cues than in the presence of familiar cues on the Stage 2 irrelevant dimension, $F(1,36) = 8.19$, $p < .01$. This difference persisted when analysis-of-covariance adjustments were made to control for differences among training level groups in Stage 2 learning, $F(1,41) = 7.83$, $p < .01$. A significant interaction between novelty and training level was obtained in the analysis of Stage 3 choice trials, $F(2,36) = 3.58$, $p < .05$. Comparisons of novelty groups within training levels indicated that a

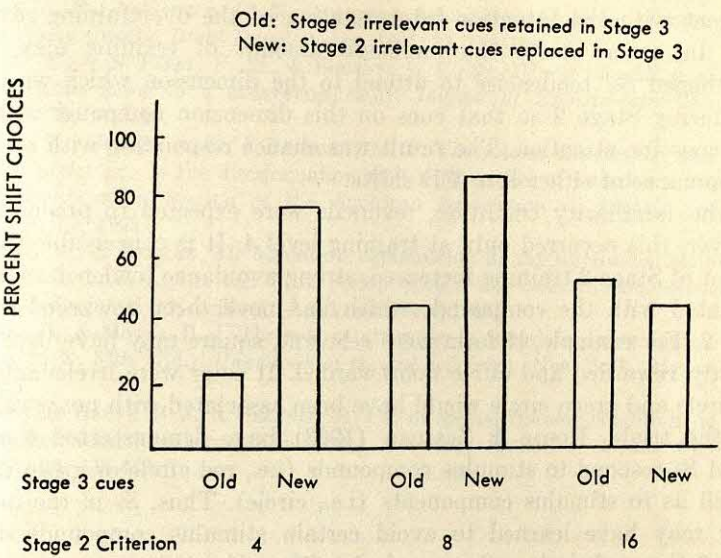


FIG. 1. Illustration of ED shift choices on Stage 3 test trials.

significant novelty effect occurred only within training levels 4 and 8 where nearly all *S*s in the novelty conditions made ED shifts, $F(1,36) = 4.57$, $p < .05$; $F(1,36) = 5.44$, $p < .05$. At training level 16, both the novelty and familiarity groups made ED shifts no more frequently than reversals, $F(1,36) = .24$, $p > .05$. Figure 1 illustrates the nature of this interaction.

DISCUSSION

The finding that *S*s in the novelty condition made more ED shifts than *S*s in the familiarity condition is consistent with an earlier finding reported by Heal, George, & Bransky (1965). In conditions that were comparable to the 8-criterion groups of the present study, they found that 62% of their *S*s made ED shifts in the presence of novel cues on a preferred dimension. However, only 22% of their *S*s made ED shifts in the presence of familiar cues on a preferred dimension. These findings suggest that cue novelty exerts an important influence on *S*'s attention to abstract properties of stimuli. Other studies have reaffirmed the tendency for novelty to influence attention in terms of increased fixation times (J. Cantor & G. Cantor, 1964a, 1964b), faster response speeds (Bogartz & Witte, 1966; G. Cantor & J. Cantor, 1964, 1965, 1966; G. Cantor & Fenson, 1968), and facilitated learning (Guy, Van Fleet, & Bourne, 1966; Zeaman, House, & Orlando, 1958).

The novelty effect was not present at training level 16. This finding is

consistent with the attention interpretation of the overtraining reversal effect. In other words, the increased amount of training may have strengthened *Ss*' tendencies to attend to the dimension which was relevant during Stage 2 so that cues on this dimension competed with the novel cues for attention. The result was chance responding with no clear predominance of either R or ED shifts.

In the familiarity condition, reversals were expected to predominate. However, this occurred only at training level 4. It is conceivable that as amount of Stage 2 training increased, strong avoidance tendencies became associated with the compounds which had never been rewarded during Stage 2. For example, if form were relevant, square may have been consistently rewarded and circle nonrewarded. If color were irrelevant both red circle and green circle would have been associated with nonreward on half the trials. House & Zeaman (1963) have demonstrated that retarded *Ss* respond to stimulus compounds (i.e., red circle or green circle) as well as to stimulus components (i.e., circle). Thus, *Ss* in the present study may have learned to avoid certain stimulus compounds during Stage 2 (i.e., red circle and green circle). Thus, if red circle were rewarded on Stage 3 training trials, *S* might have retained the tendency to avoid green circle. This tendency may have competed with the new tendency to approach circle and avoid square, resulting in chance performance on choice trials. The overall results of this study emphasize the central role of attention in discrimination learning. These findings further suggest the possibility that learning deficits may be overcome by the use of procedures which influence the direction and duration of attention.

REFERENCES

Bogartz, R., & Witte, K. L. On the locus of the stimulus familiarization effect in young children. *Journal of Experimental Child Psychology*, 1966, **4**, 317–331.

Cantor, G. N., & Cantor, J. H. Effects of conditioned-stimulus familiarization on instrumental learning in children. *Journal of Experimental Child Psychology*, 1964, **1**, 71–78.

Cantor, G. N., & Cantor, J. H. Discriminative reaction time performance in preschool children as related to stimulus familiarization. *Journal of Experimental Child Psychology*, 1965, **2**, 1–9.

Cantor, G. N., & Cantor, J. H. Discriminative reaction time in children as related to amount of stimulus familiarization. *Journal of Experimental Child Psychology*, 1966, **4**, 150–157.

Cantor, G. N., & Fenson, L. New data regarding the relationship of amount of familiarization to the stimulus familiarization effect (SFE). *Journal of Experimental Child Psychology*, 1968, **6**, 167–173.

Cantor, J. H., & Cantor, G. N. Children's observing behavior as related to amount and recency of stimulus familiarization. *Journal of Experimental Child Psychology*, 1964, **1**, 241–247. (a)

CANTOR, J. H., & CANTOR, G. N. Observing behavior in children as a function of stimulus novelty. *Child Development,* 1964, **35,** 119–128. (b)

GUY, D. E., VAN FLEET, F. M., & BOURNE, L. E., JR. Effects of adding a stimulus dimension prior to a nonreversal shift. *Journal of Experimental Psychology,* 1966, **72,** 161–168.

HEAL, L. W., GEORGE, R. T., & BRANSKY, M. L. The role of cue novelty and dimension preference in the discrimination shift performance of retardates. Paper read at Southeastern Section of the American Association on Mental Deficiency, Nashville, 1965.

HOUSE, B. J., & ZEAMAN, D. Miniature experiments in the discrimination of retardates. In L. P. Lipsitt and C. C. Spiker (Eds.), *Advances in child development and behavior* (I). New York: Academic Press, 1963.

ZEAMAN, D., & HOUSE, B. J. The role of attention in retardate discrimination learning. In N. Ellis (Ed.), *Handbook of mental deficiency.* New York: McGraw-Hill, 1963.

ZEAMAN, D., HOUSE, B. J., & ORLANDO, R. Use of special training conditions in visual discrimination learning with imbeciles. *American Journal of Mental Deficiency,* 1958, **63,** 453–459.

Editorial Note

As the number of manuscripts submitted to *Journal of Experimental Child Psychology* increases, the Editor and the Editorial Board have become increasingly dependent on our colleagues to serve as guest editors so that there will be at least two reviewers for each paper submitted. During the past year, 1969, the following psychologists have helped us with the reviewing process. "Helped us" does not adequately describe what they did, for in fact, each has set aside time to review thoroughly and constructively the manuscript assigned. Contributors of manuscripts have on many occasions written to express their gratitude for the unusually conscientious assistance of the reviewers. I join them.

SIDNEY W. BIJOU
Editor

Herbert G. Birch, *Albert Einstein College of Medicine*
Erling E. Boe, *University of Pennsylvania*
Jay S. Birnbrauer, *University of North Carolina*
William H. Bogartz, *University of California*
Lila Ghent Braine, *University of California*
Martin D. Braine, *University of California*
Jerome S. Bruner, *Harvard University*
Joseph C. Campione, *University of Washington*
Joan H. Cantor, *University of Iowa*
Rachel K. Clifton, *University of Massachusetts*
Leslie B. Cohen, *University of Illinois*
Virginia C. Crandall, *Fels Research Institute*
Bruce R. Elkstrand, *University of Colorado*
David Elkind, *University of Rochester*
Barbara C. Etzel, *University of Kansas*
Susan Ervin-Tripp, *University of California*
J. H. Flavell, *University of Minnesota*
LeRoy H. Ford, *State University at New York*
Frank A. Geldard, *551 Lake Dr., Princeton, N. J.*
Rochel S. Gelman, *University of Pennsylvania*
Jacob L. Gewirtz, *National Institute of Mental Health*
Joseph A. Glick, *University of Minnesota*
John W. Gyr, *University of Michigan*

Marshall M. Haith, *Harvard University*
Lauren Harris, *Michigan State University*
Robert P. Hawkins, *Western Michigan University*
Maurice E. Hershenson, *University of Wisconsin*
Kennedy T. Hill, *University of Illinois*
James G. Holland, *University of Pittsburgh*
Billy L. Hopkins, *Southern Illinois University*
Wendell E. Jeffrey, *University of California*
Geoffrey Keppel, *University of California*
William Kessen, *Yale University*
Marvin Levine, *State University of New York*
Lewis P. Lipsitt, *Brown University*
Norman H. Mackworth, *Stanford University*
Charles H. Madsen, *Florida State University*
Irving M. Maltzman, *University of California*
William R. Mead, *University of Illinois*
William J. Meyers, *University of Iowa*
Norman A. Milgram, *Temple University*
William E. Montague, *University of Illinois*
Mona Morningstar, *Stanford University*
Peter Murdoch, *Smith Richardson Foundation*
Nancy A. Myers, *University of Massachusetts*
Donald A. Norman, *University of California*
Daniel O'Leary, *State University of New York*
John J. O'Neill, *University of Illinois*
Robert Orlando, *Plymouth Home & Training School*
Sonia F. Osler, *Adelphi University*
Allen U. Paivio, *University of Western Ontario*
David A. Parton, *University of Iowa*
Ronald K. Penney, *University of Waterloo*
Robert F. Peterson, *University of Illinois*
Michael I. Posner, *University of Oregon*
F. Michael Rabinowitz, *University of Washington*
V. Edward Renner, *University of Illinois*
Frank J. Restle, *Indiana University*
Morton Rieber, *San Francisco State College*
M. E. Rosenbaum, *University of Iowa*
Thomas J. Ryan, *Carleton University*
Rudolph W. Schulz, *University of Iowa*
Keith G. Scott, *University of Illinois*
James Sherman, *University of Kansas*
Howard N. Sloane, Jr., *University of Utah*

Janet T. Spence, *University of Texas*
Charles C. Spiker, *University of Iowa*
Robert L. Sprague, *University of Illinois*
Theodore E. Steiner, *University of Illinois*
Warren M. Steinman, *University of Illinois*
Lawrence T. Stoddard, *Massachusetts General Hospital*
Louise S. Tighe, *Stanford University*
Endel Tulving, *University of Toronto*
Robert A. Wahler, *University of Tennessee*
Herbert H. Wells, *University of North Carolina*
Morton M. Weir, *University of Illinois*
Grover J. Whitehurst, *University of Illinois*
Joseph L. Wolff, *University of Illinois*
Michael D. Zeiler, *University of Iowa*
Edward F. Zigler, *Yale University*

Announcement

In order to achieve greater standardization of style within journals of psychology, authors are requested to adhere closely in all details to the APA Style Manual. Your attention is called specifically to the appropriate use of ampersands rather than the word "and" in the text and in the reference lists where there is more than one author's name mentioned. Furthermore, authors are requested to use the metric system whenever possible.

AUTHOR INDEX

A

Ainsfeld, Moshe, 250

B

Bernstein, Martin E., 63
Bilsky, Linda, 503
Birch, Herbert G., 483
Brethower, Dale, 45
Briggs, Peter F., 356
Bryant, Brenda, 250

C

Cabot, Raymond, 45
Cantor, Gordon N., 74
Cheyne, J. Allan, 127
Clarke, Ann M., 180
Cooper, G. M., 180

D

Davis, Linda, 483
Di Vesta, Francis J., 63
Dodd, Cornelia, 296

E

Eckerman, Carol O., 271
Eichberg, Robert, 195
Eimas, Peter D., 160

F

Fagan, Joseph F., III, 9
Fouts, Gregory T., 118, 461
Frobish, M. J., 40

G

Gaines, Rosslyn, 418
Gallagher, Joseph W., 411
Gallimore, Ronald, 140
Gavalas, Rochelle, 356
Goss, Albert E., 402
Gotway, Mary Ann, 190
Goyeche, John R. M., 127

H

Halford, Graeme S., 314
Heal, Laird W., 503
Herron, R. E., 40
Hooper, Frank H., 234
Horowitz, Leonard M., 375

I

Immergluck, Ludwig, 210

J

Jeffrey, Wendell E., 256

K

Kausler, Donald H., 190
Kelley, Crystal A., 306
Kemp, Bryan, 140
Kennedy, Dale B., 222
Kerr, Nancy J., 147
Kling, J. Kenneth, 82
Kozma, Albert, 1
Kubose, Sunnan K., 74

L

Lampel, Anita K., 375
Lewis, Michael, 296
Lipsitt, Lewis P., 263
Loudon, Elizabeth H., 180
Lumsden, Ernest A., 82

M

Mahoney, Kurt, 147
Mearini, Maria Claudia, 210
Meyerson, Lee, 147
Miller, Frank D., 106
Mithaug, Dennis E., 443
Modreski, Regina A., 402
Moffett, Adrienne, 173
Moreau, Tina, 000

N

Naito, Toru, 263
Nelson, Katherine J., 256, 284

P

Parton, David A., 118, 461
Pascual-Leone, Juan, 328
Reingold, Harriet L., 271

R

Ross, Bruce M., 20
Rovee, Carolyn Kent, 33
Rovee, David T., 33

S

Scott, Keith G., 432
Shepp, Bryan E., 389
Smith, Donald E. P., 45
Smith, June, 328
Speer, David C., 356
Stave, Mary, 195

T

Takanishi, Ruby N., 375
Tharp, Roland G., 140
Tighe, Louise S., 366, 494
Tighe, Thomas J., 366, 494
Trabasso, Tom, 195
Turkewitz, Gerald, 483
Turrisi, Frank D., 389

U

Urbano, Richard C., 432

V

Van Wagenen, R. Keith, 147

W

Walters, Richard H., 127
Watson, Peter, 92
Weisberg, Paul, 222
Wickelgren, Lyn W., 469

Y

Youniss, James, 20

Z

Zeiler, Michael D., 306
Zelniker, Tamar, 256

SUBJECT INDEX

A

Age trends, concept formation, 256
color and form naming, 402
conservation, 82, 234
discrimination, 160
free recall, 284
memory, 20
paired-associate learning, 63, 190
perception, 210
symbolic behavior, 328, 411
visual perception, 418
Alternation, 92
Anxiety, 1, 127
Attention, 127, 140, 173

C

Computer analysis, 40
Concept formation, 256
Conditioning, eyelid, 263
Conservation, 82, 234, 314
Cooperation, 443

D

Dependency behavior, 356
Discrimination, form, 418
learning, 160, 195, 366, 389
matching, 461
mediation, 366
novelty effects, 503
of size, 82
pretraining, 494
reversal, 366, 389, 494
transfer, 180, 195, 494, 503

E

Eyemovements, infant, 469
Exploratory behavior, 33
infant, 271

F

Familiarization, 74, 106

I

Imitation, 118
Infants, 33
conditioning, 263
exploration, 271
eyemovement, 469
perception, 173
somesthesis, 483
Intelligence, effects on digit span, 432

L

Language, kindergarten testing, 250
learning, 411
semantics, 411
Learning, alternation, 92
free recall, 9, 284
paired-associate, 63
R-S, 190

M

Matching, 1, 461
Memory, organization, 375
recognition, 20
retardate, 432
Movement analysis, 40

O

Orienting response, 296

P

Paired-Associates, 63, 190
Perception, form, 418
visual, 74, 173, 210
Preferences, form and color, 195
Punishment, 127

R

Reaction time, 106
Reading, 45
Recall, 9
Reinforcement, conjugate, 33
Reinforcement, of classroom behavior, 45
of cooperation, 443
partial, 92
schedules, 222, 306, 356
social, 1, 140, 356
Retardates, 9, 45, 147
digit span, 432
discrimination, 503

discrimination learning, 389
Retention, 9, 20, 375
organization, 284

S

Social isolation, 1
Somesthesis, infants, 483
Sex differences, conversion, 234
perception, 210
social behavior, 356
Symbolic behavior, 328

T

Toilet training, 147
Transposition, 366

JOURNAL OF Experimental Child Psychology

EDITOR: Sidney W. Bijou

ASSOCIATE EDITORS: Richard S. Bogartz
David Zeaman

EDITORIAL BOARD:

Albert Bandura
Wesley C. Becker
Harry Beilin
Lyle E. Bourne, Jr.
Gordon N. Cantor
Charles C. Clifton, Jr.
Peter Eimas
Robert L. Fantz
Eleanor J. Gibson
Eugene S. Gollin
Willard W. Hartup
Raymond H. Hohle
Herbert Kaye
Tracy S. Kendler
Helen L. Koch
Eugene R. Long
Harry Munsinger
David S. Palermo
Ross Parke
Herbert L. Pick
Hayne W. Reese
Todd Risley
Bryan E. Shepp
Murray Sidman
Einar Siqueland
Patrick Suppes
Thomas M. Trabasso
Paul Weisberg
Joachim Wohlwill
Betty House Zeaman

Volume 8, 1969

ACADEMIC PRESS
New York and London

CONTENTS OF VOLUME 8

NUMBER 1, AUGUST 1969

ALBERT KOZMA. The Effects of Anxiety, Stimulation, and Isolation on Social Reinforcer Effectiveness 1

JOSEPH F. FAGAN III. Free Recall Learning in Normal and Retarded Children 9

BRUCE M. ROSS AND JAMES YOUNISS. Ordering of Nonverbal Items in Children's Recognition Memory 20

CAROLYN KENT ROVEE AND DAVID T. ROVEE. Conjugate Reinforcement of Infant Exploratory Behavior 33

R. E. HERRON AND M. J. FROBISH. Computer Analysis and Display of Movement Patterns 40

DONALD E. P. SMITH, DALE BRETHOWER, AND RAYMOND CABOT. Increasing Task Behavior in a Language Arts Program by Providing Reinforcement 45

FRANCIS J. DI VESTA AND MARTIN E. BERNSTEIN. The Effect of Serial Structure on Children's and Adults' Paired-Associate Learning 63

GORDON N. CANTOR AND SUNNAN K. KUBOSE. Preschool Children's Ratings of Familiarized and Nonfamiliarized Visual Stimuli . 74

ERNEST A. LUMSDEN AND J. KENNETH KLING. The Relevance of an Adequate Concept of "Bigger" for Investigations of Size Conservation: A Methodological Critique 82

PETER WATSON. Partial Reward and Double Alternation Learning in Children 92

FRANK D. MILLER. The Effects of Differential Amounts of Stimulus Familiarization on Choice Reaction Time Performance in Children 106

GREGORY T. FOUTS AND DAVID A. PARTON. Imitation: Effects of Movement and Static Events 118

J. ALLAN CHEYNE, JOHN R. M. GOYECHE, AND RICHARD H. WALTERS. Attention, Anxiety, and Rules in Resistance-To-Deviation in Children 127

RONALD GALLIMORE, ROLAND G. THARP, AND BRYAN KEMP. Positive Reinforcing Function of "Negative Attention" 140

R. KEITH VAN WAGENEN, LEE MEYERSON, NANCY J. KERR, AND KURT MAHONEY. Field Trials of a New Procedure for Toilet Training 147

PETER D. EIMAS. A Developmental Study of Hypothesis Behavior and Focusing 160

ADRIENNE MOFFETT. Stimulus Complexity as a Determinant of Visual Attention in Infants 173
ANN M. CLARKE, G. M. COOPER, AND ELIZABETH H. LOUDON. A Set to Establish Equivalence Relations in Pre-school Children . 180
DONALD H. KAUSLER AND MARY ANN GOTWAY. R-S Learning in Children 190

NUMBER 2, OCTOBER 1969

TOM TRABASSO, MARY STAVE, AND ROBERT EICHBERG. Attitude Preference and Discrimination Shifts in Young Children . . 195
LUDWIG IMMERGLUCK AND MARIA CLAUDIA MEARINI. Age and Sex Differences in Responses to Embedded Figures and Reversible Figures 210
PAUL WEISBERG AND DALE B. KENNEDY. Maintenance of Children's Behavior by Accidental Schedules of Reinforcement . . . 222
FRANK H. HOOPER. Piaget's Conservation Tasks: The Logical and Developmental Priority of Identity Conservation 234
BRENDA BRYANT AND MOSHE AINSFELD. Feedback versus No-Feedback in Testing Children's Knowledge of English Pluralization Rules 250
KATHERINE J. NELSON, TAMAR ZELNIKER, AND WENDELL E. JEFFREY. The Child's Concept of Proportionality: A Re-examination . 256
TORU NAITO AND LEWIS P. LIPSITT. Two Attempts to Condition Eyelid Responses in Human Infants 263
HARRIET L. RHEINGOLD AND CAROL O. ECKERMAN. The Infant's Free Entry into a New Environment 271
KATHERINE J. NELSON. The Organization of Free Recall by Young Children 284
CORNELIA DODD AND MICHAEL LEWIS. The Magnitude of the Orienting Response in Children as a Function of Changes in Color and Contour 296
MICHAEL D. ZEILER AND CRYSTAL A. KELLEY. Fixed-Ratio and Fixed-Interval Schedules of Cartoon Presentation 306
GRAEME S. HALFORD. An Experimental Analysis of the Criteria Used by Children to Judge Quantities 314
JUAN PASCUAL-LEONE AND JUNE SMITH. The Encoding and Decoding of Symbols by Children: A New Experimental Paradigm and a Neo-Piagetian Model 328
DAVID C. SPEER, PETER F. BRIGGS, AND ROCHELLE GAVALAS. Concurrent Schedules of Social Reinforcement and Dependency Behavior among Four-Year-Old Children 356

THOMAS J. TIGHE AND LOUISE S. TIGHE. Facilitation of Transposition and Reversal Learning in Children by Prior Perceptual Training 366
LEONARD M. HOROWITZ, ANITA K. LAMPEL, AND RUBY N. TAKANISHI. The Child's Memory for Unitized Scenes 375
FRANK D. TURRISI AND BRYAN E. SHEPP. Some Effects of Novelty and Overtraining on the Reversal Learning of Retardates . 389
REGINA A. MODRESKI AND ALBERT E. GOSS. Young Children's Initial and Changed Names for Form-Color Stimuli . . . 402
ERRATA 410

NUMBER 3, DECEMBER 1969

JOSEPH W. GALLAGHER. Semantic Relationship and the Learning of Syntactic Word Pairs in Children 411
ROSSLYN GAINES. The Discriminability of Form Among Young Children 418
RICHARD C. URBANO AND KEITH G. SCOTT. Digit Span, Practice and Dichotic Listening Performance in the Mentally Retarded 432
DENNIS E. MITHAUG. The Development of Cooperation in Alternative Task Situations 443
DAVID A. PARTON AND GREGORY T. FOUTS. Effects of Stimulus-Response Similarity and Dissimilarity on Children's Matching Performance 461
LYN W. WICKELGREN. The Ocular Response of Human Newborns to Intermittent Visual Movement 469
GERALD TURKEWITZ, TINA MOREAU, LINDA DAVIS, AND HERBERT G. BIRCH. Factors Affecting Lateral Differentiation in the Human Newborn 483
LOUISE S. TIGHE AND THOMAS J. TIGHE. Transfer from Perceptual Pretraining as a Function of Number of Task Dimensions . 494
LINDA BILSKY AND LAIRD W. HEAL. Cue Novelty and Training Level in the Discrimination Shift Performance of Retardates 503

EDITORIAL NOTE 512

ANNOUNCEMENT 515

AUTHOR INDEX 516

SUBJECT INDEX 518

INFORMATION FOR AUTHORS

The *Journal of Experimental Child Psychology* will publish papers in which the behavior and development of children is clearly related to its determining variables. Typically (not always) this would mean that a variable has been manipulated in an experimental manner. Articles will be reviewed in terms of the apparent reliability of their findings and the clarity of the relationships they affirm. Research conceived in any theoretical framework is invited, so long as its conclusions are reliable and compelling. Occasionally, critical reviews of related bodies of experimental literature, theoretical discussions which define new paths of experimental work, and well-controlled developmental studies on infra-human subjects will be published. The criterion for acceptance of such papers will be their value in prompting experimental research in related areas with children. Short notes describing new apparatus or techniques, new materials, circuits, or other information valuable to investigators in the field will be accepted from time to time. There is no page charge for publication.

Two copies of each manuscript and each figure or table should be sent to:

Dr. Sidney W. Bijou, Child Behavior Laboratory, 403 E. Healey, University of Illinois, Champaign, Illinois 61820

A third copy should be retained by the author. Manuscripts must be double-spaced and may be typed, dittoed, or mimeographed. Each manuscript must include a concise abstract, preferably not exceeding 300 words, which will appear after the title and author's name in the published article. In some papers this abstract will obviate the need for a summary. In general, authors are requested to follow the specifications of the *Publication Manual* of the American Psychological Association.

Figures and Tables. Figures should be numbered consecutively with Arabic numerals in order of mention in the text; each figure should have a descriptive legend. Legends should be typed together on a separate sheet, double-spaced. All illustrations should be in finished form ready for the engraver. A convenient size for drawings is $8\frac{1}{2} \times 11$ inches (21×27.5 cm). Drawings should be made with India ink on tracing linen, smooth surface white paper, or bristol board. Graph paper if used should be ruled in blue. Grid lines that are to show in the final engraving should be inked in black. Sharp, clear glossy prints of figures are acceptable.

Tables should be typed on separate pages, numbered consecutively with Arabic numerals in order of mention in the text. All tables should have descriptive headings, typed (double-spaced) above the table.

Literature references in the text should be cited by author's surname and the year of publication, e.g., Smith, 1960; Smith *et al.*, 1959; Smith and Jones, 1961, p. 250 (for references to a specific page); (Smith & Jones, 1962) (ampersand for references in parentheses). If a reference has more than two authors, the citation includes the surnames of all authors at the first mention, but later citations of the same reference include only the surname of the first author and the abbreviations *"et al."* Suffixes *a, b,* etc., should be used following the date to distinguish two or more works by the same author(s) in the same year, e.g., Smith, 1960a, 1960b. Literature citations in the bibliography should be arranged alphabetically according to the surname of the author.

Journal titles should be written out in full according to the form followed in the Psychological Abstracts (Vol. 39).

The desired style of capitalization and punctuation in references is apparent in the following examples:

STAATS, A. W., & STAATS, C. K. A comparison of the development of speech and reading behavior with implications for research. *Child Development,* 1962, **33**, 831–846.

SPIKER, C. C. Research methods in children's learning. In P. H. Mussen (Ed.), *Handbook for research methods in child development.* New York: Wiley, 1960. Pp. 374–420.

MUNN, N. L. *The evolution and growth of human behavior.* Cambridge: Riverside Press, 1955.

Reprints: Fifty gratis reprints are supplied and additional copies may be ordered.

A Volume in **SOCIAL PSYCHOLOGY**
A Series of Monographs, Treatises and Texts
Edited by **Leon Festinger** and **Stanley Schachter**

EXPERIMENTS IN SOCIAL PSYCHOLOGY

By **Paul G. Swingle**
McGill University

Stressing the development of research skills in experimental social psychology, this text exposes the student to a variety of research areas and problems of methodology including observer bias, brainstorming, conformity, psychological gaming, and attitude change.

Designed for use in experimentally oriented undergraduate courses in conjunction with a basic social psychology text, this volume provides stimulating experimental research projects requiring no specialized equipment. Statistical study is not a prerequisite, since the student is encouraged to plot, examine and understand his data rather than concern himself with establishing the reliability of differences.

A comprehensive introduction precedes each project, and reprinted articles provide theoretical background.

1969, 292 pp., $7.50

AP 2332

Academic Press
NEW YORK AND LONDON
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
BERKELEY SQUARE HOUSE, LONDON W.1